Technical Report: NAVTRAEQUIPCEN 73-C-0128-1

DECISION MAKING AND TRAINING:
A Review of Theoretical and Empirical Studies of Decision Making and Their Implications for the Training of Decision Makers

R. S. Nickerson
C. E. Feehrer

Bolt Beranek and Newman, Inc.
Cambridge, Massachusetts 02138
Contract N61339-73-C-0128 new

12

August 1975

DOD Distribution Statement

Approved for public release;
distribution unlimited.

NAVAL TRAINING EQUIPMENT CENTER
ORLANDO, FLORIDA 32813

ABSTRACT

This report reviews theoretical and empirical studies of decision making. The purpose of the review was to identify results that would be applicable to the problem of training decision makers.

The literature on decision making is extensive. However, relatively few studies have dealt explicitly with the problem of training in decision-making skills. The task, therefore, was to gather from the general literature on decision making any implications that could be found for training.

Decision making is conceptualized here as a type of problem solving, and the review is organized in terms of the following component tasks: information gathering, data evaluation, problem structuring, hypothesis generation, hypothesis evaluation, preference specification, action selection, and decision evaluation. Implications of research findings for training are discussed in the context of descriptions of each of these tasks.

A general conclusion drawn from the study is that decision making is probably not sufficiently well understood to permit the design of an effective general-purpose training system for decision makers. Systems and programs could be developed, however, to facilitate training with respect to specific decision-making skills. The development of more generally applicable training techniques or systems should proceed in an evolutionary fashion.

Training is one way to improve decision-making performance; another is to provide the decision maker with aids for various aspects of his task. Because training and the provision of decision aids are viewed as complementary approaches to the same problem, the report ends with a discussion of several decision-aiding techniques that are in one or another stage of study or development.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
NAVTRAEQUIPC[redacted]73-C-0128-1

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Decision Making and Training: A Review of Theoretical and Empirical Studies of Decision Making and their Implications for the Training of Decision Makers.

5. TYPE OF REPORT & PERIOD COVERED
Research 6/73 - 7/75

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
R. S. NICKERSON
C. E. FEEHRER

8. CONTRACT OR GRANT NUMBER(s)
~~NAVTRAEQUIPCEN~~
N61339-73-C-0128[redacted]

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Bolt Beranek and Newman, Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NAVTRAEQUIPCEN
Task No. 3751-01[redacted]

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Training Equipment Center
Code N-215
Orlando, Florida 32813

12. REPORT DATE
Aug[redacted]75

13. NUMBER OF PAGES
210

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision making	Preference Specification	Training
Problem solving	Action Selection	Training Device Technology
Information gathering	Decision Aids	
Hypothesis Generation	Probability theory	
Hypothesis Evaluation	Command and Control	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report reviews theoretical and empirical studies of decision making. The purpose of the review was to identify results that would be applicable to the problem of training decision makers.

The literature on decision making is extensive. However, relatively few studies have dealt explicitly with the problem of training in decision-making skills. The task, therefore, was to gather from the general literature on decision making any implications that could be found for training. (Cont)

Decision making is conceptualized here as a type of problem solving, and the review is organized in terms of the following component tasks: information gathering, data evaluation, problem structuring, hypothesis generation, hypothesis evaluation, preference specification, action selection, and decision evaluation. Implications of research findings for training are discussed in the context of descriptions of each of these tasks.

A general conclusion drawn from the study is that decision making is probably not sufficiently well understood to permit the design of an effective general-purpose training system for decision makers. Systems and programs could be developed, however, to facilitate training with respect to specific decision-making skills. The development of more generally applicable training techniques or systems should proceed in an evolutionary fashion.

Training is one way to improve decision-making performance; another is to provide the decision maker with aids for various aspects of his task. Because training and the provision of decision aids are viewed as complementary approaches to the same problem, the report ends with a discussion of several decision-aiding techniques that are in one or another stage of study or development.

FOREWORD

The Human Factors Laboratory of the Naval Training Equipment Center has been involved in decision-making research with the objective of developing an approach to decision-making training which will improve the decision-making and tactical performance capabilities of Navy commanders. This report is the result of an analytical review of decision-making research which was performed to identify information pertinent to the training of decision-making skills.

The outcome of this effort corroborated an impression that very little of the great amount of decision-making research has directly addressed the problem of training in decision making. The review has identified implications for the training of decision makers and areas for research which could provide insight for the development of effective training procedures and programs.

William P. Lane

WILLIAM P. LANE
Acquisition Director

TABLE OF CONTENTS

PRECEDING PAGE BLANK-NOT FILMED

List of Tables

List of Illustrations

List of Illustrations (Cont)

SECTION I

INTRODUCTION

Much has been written about the importance of decision making for industry, for government, for the military and for rational--or at least reasonable--people in general. Moreover, a great deal of research has been conducted on decision-making behavior. In spite of these facts--or perhaps because of them--there is not general agreement concerning what decision making is, how it should be done, how it is done, how to tell whether it is done well or poorly, and how to train people to do it better.

The term "decision making" has been applied to a very broad range of behaviors. The detection of weak sensory stimuli has been viewed, in part, as a decision process (Green & Swets, 1966), as has perception by humans more generally (Bruner, 1957). Pattern classification by machines (Sebestyn, 1962), the retrieval of information from memory (Egan, 1958), the performance of skilled tasks such as automobile driving (Algea, 1964) and airplane piloting (Szafran, 1970), the production of speech (Rochester & Gill, 1973), educational counseling (Stewart & Winborn, 1973), the purchasing of industrial products (Reingen, 1973), the evaluation of the performance of salesmen (Sheridan & Carlson, 1972), and the conducting of a laboratory experiment (Edwards, 1956) are also representative of the types of processes that have been discussed under the rubric of decision making. Probably when the term is used in industrial, governmental and military contexts, however, what the user has in mind is something close to what Schrenk (1969) describes as "situations characterized by fairly well-defined objectives, significant action alternatives, relatively high stakes, inconclusive information and limited time for decision" (p. 544). We hasten to add that to limit one's attention to situations that have all of these characteristics would preclude consideration of the large majority of experimental investigations of decision making; in particular, in very few laboratory studies of decision making have the stakes been high; and one may question in many--if not most--cases the significance of the action alternatives to the experimental subjects. It does not necessarily follow that the results of laboratory studies have no relevance to real-life decision making, of course. The degree to which one is willing to extrapolate from the one situation to the other depends on the extent to which one subscribes to the view that simple and inconsequential decision problems are solved--at least in principle --in the same ways as are those that are complex and consequential.

As Schrenk (1969) has pointed out, there are three ways to improve the performance of the human decision element in a system: (1) selection (insure that decisions are made only by individuals who are competent to make them), (2) training (attempt to improve the decision-related skills of people in decision-making positions),

and (3) decision aiding (provide decision makers with procedural and technical aids to compensate for their own limitations). To the extent that performance of a decision-making system is of interest, as opposed to that of a human being, another possibility that deserves consideration is that of automation (have machines perform those decision tasks that they can perform better than people).

The number of tasks that are now performed by machines that were once thought to require human skills is growing and will continue to do so. Many tasks that involve decision making by some definition should be--indeed, many have been--automated. There is little justification for wasting a good human brain to make what Soelberg (1967) calls "programmed decisions," decisions that are made with sufficient frequency and under sufficiently specifiable conditions to permit the detailed description of procedures for making them. Thermostats, governors, regulators, stabilizers, computer algorithms, and such things, are the preferred "decision makers" for these types of situations. The situations with which we are primarily concerned are not of this straightforward programmed type. They are situations that are novel, unstructured or unplanned for, or they involve human preferences that are not easily specified, or potential action consequences that are not known with certainty. Clearly, these types of situations are the more interesting objects of study, and are probably more representative of what people view as bona fide decision making.

It is important to recognize that the objectives of much decision-making research are to make novel situations less novel by providing prototypes in terms of which the novel situations can be perceived, to facilitate the imposition of structure on situations when apparent structure is lacking, and to provide techniques for decreasing the probability of surprises and for coping with unplanned-for situations as though they had been anticipated all along. But the reader who might think that such objectives could, if realized, take the charm out of decision making may rest easy. There seems little danger of success to the point of reducing all decision making to an algorithmic process in the near future. Indeed, there are some aspects of decision making that men may never feel comfortable turning over to machines. Hence, the needs for selection, training and decision aiding are still real, and are likely to continue to be for some time to come. Moreover, as more and more of the procedurizable tasks that were once performed by men do become automated, the tasks that are left to be performed by men--or perhaps by men and machines in collaboration--take on added interest and significance by virtue of their very resistance to automation. Should not those tasks which seem to require the attention of human brains be the tasks that hold a unique fascination for us as human beings?

The general question that motivates this study is the question of whether individuals can be trained to be effective decision makers in unprogrammed situations. And if the answer to that question appears to be yes, the next question that presents itself is that of how that training can be accomplished most effectively. Immediately, one is led to more specific questions. Does it make sense to think of decision making as a skill, or as a collection of skills, that can be developed in a sufficiently general way that they can be applied in a variety of specific contexts? What is it that the decision maker needs to be taught? Concepts? Facts? Principles? Attitudes? Procedures? Heuristics?

The literature on decision-making research is volumious, but despite numerous references to the importance of the training of decision makers (e.g., Edwards, 1962; Evans & Cody, 1969; Fleming, 1970; Hammell & Mara, 1970; Kanarick, 1969; Kepner & Tregoe, 1965; Scalzi, 1970; Sidorsky & Simoneau, 1970), the number of studies that have explicitly addressed the question of exactly what should be taught and how the teaching can best be accomplished is remarkably small. The central interest in the area continues to be with parameterization of the decision maker and his environment and with generation of specific aids to the decision process.

This review is not limited, therefore, to studies that have focused specifically on the issue of decision training. We have attempted instead to look at a rather broad cross section of the general decision-making research literature with a view to finding, wherever we could, implications for the training of decision makers and clues concerning what further research might lead to more effective training procedures or programs.

SECTION II

SOME COMMENTS ON DECISION THEORY

One can distinguish two rather different approaches that have been taken to the study of decision making. One is analytical; the other is basically empirical. A common goal of both approaches, however, is the development of formal models of decision processes. In the first case, one tries to analyze decision situations--often hypothetical situations--abstracting from them their common elements. One then attempts to produce a model of the decision-making process, using the constructs that have been identified in the process of analysis. In the empirical approach, one begins by observing individuals making decisions in real-life situations, and attempts, on the basis of these observations, to develop parsimonious descriptions of decision-making behavior.

Each approach has its strengths and its weaknesses. The models generated by analysis are likely to be more abstract than those developed through observation. As a consequence, they are typically more general. However, there may be considerable difficulty in applying such models in specific cases. This is true because real-life decision situations frequently are not easily describable in terms that an application of a model would require. In contrast, a model of a decision-making process that is developed by observing decision makers in action is likely to be applicable, at least to situations highly similar to that from which the model is derived. Such models may lack generality, however, and prove to be inapplicable outside the context in which they are developed.

2.1 Prescriptive versus Descriptive Models

A prescriptive model indicates what one should do in a given decision situation; a descriptive model is intended to describe what one actually does. Typically, prescriptive models are the outcomes of analytical approaches to the study of decision making, whereas empirical approaches generally lead to descriptive models. In theory at least, a prescriptive model may be used either as a guide for decision makers or as a standard against which to assess the extent to which decision-making performance approaches optimality. Descriptive models differ from prescriptive models insofar as human decision makers perform in a less than optimal fashions. Were a decision maker to behave in an optimal fashion, a description of his behavior would constitute a prescriptive model. Comparisons between prescriptive and descriptive models can be instructive in suggesting the reasons why human behavior is sometimes not optimal.

Prescriptive models are generally associated with economists and mathematical statisticians. Among the developers and expositors

of prescriptive decision theory are Bernoulli (1738), Neyman and Pearson (1933), Samuelson (1947), vonNeumann and Morgenstern (1947), Wald (1947, 1950), Good (1952), Blackwell and Girshick (1954), Savage (1954), Luce and Raiffa (1957), and Schlaifer (1959). Such models typically postulate an "economic," or at least a "rational," man who behaves in a way that is entirely consistent with his decision objectives and who does not have some of the limitations of real people.

Descriptive models were introduced primarily by psychologists and other students of human behavior, notably Edwards (1954, 1961); Peterson, Birdsall, and Fox (1954); Thrall, Coombs, and Davis (1954); Simon (1954, 1955); Tanner (1956); Davidson, Suppes, and Siegel (1957); Festinger (1957); Luce (1959); Siegel (1959); Rapoport (1960); Estes (1961); and Edwards, Lindman, and Savage (1963). The objective in this case has been to discover by experiment and observation how human beings, given their limitations, perform in decision-making situations. It is important to note that descriptive models have been viewed as descriptive only of the behavior of the decision maker, and not necessarily of the thinking that leads to that behavior. For example, the finding that an individual's choice between two gambles can be predicted on the basis of which has the most favorable "expected outcome" is not taken as evidence that in making the choice the individual actually goes through the process of calculating expected values and picking the alternative with the largest one (Edwards, 1955; Ellsberg, 1961).

The two lines of development--prescriptive and descriptive models--have not proceeded independently of each other. Several of the investigators mentioned above have made significant contributions of both prescriptive and descriptive types. Moreover, one approach that has been taken to the study of human limitations is that of attempting to modify prescriptive models so that they are in fact more descriptive. Typically, what this involves is the imposition of constraints on the model that represent specific limitations of the human. For example, a prescriptive model that assumes an infallible memory of unlimited capacity is unlikely to be very descriptive of human behavior; to modify such a model for the purpose of increasing its descriptiveness would necessitate at least the addition of some constraints that represent such factors as a limitation on memory capacity and degradation of stored information over time.

The distinction between prescriptive and descriptive models is sometimes blurred in the literature and one cannot always be sure in which way a proponent of a model intends for it to be taken. On the other hand, many writers have observed that the models deriving from theories of economics do, in fact, fail to describe behavior, or at least to do so very accurately. Miller

and Starr (1967) point out, for example, that in the economist's view of decision making, the objective of the decision maker is to maximize the "utility" that he can achieve within the limitations of his resources. They note, however, that the assumption that individuals do act so as to maximize utility has been challenged by many investigators of decision making. If rationality is defined in terms of the extent to which behavior is appropriate to the maximization of utility, they note, then when people do not maximize utility, they, by definition, are acting irrationally. Miller and Starr list several factors that have been suggested as possible reasons for the failure of decision makers to behave in an optimal way: "the inability of the individual to duplicate the rather recondite mathematics which economists have used to solve the problem of maximization of utility; the existence of other values which, though not readily quantifiable, do cause divergences from the maximization of utility in the marketplace; the effect of habit; the influence of social emulation; the effect of social institutions" (p. 25).

While interest in prescriptive models stems at least in part from the assumption that they can provide guidance for decision makers in real-life situations, their application often proves to be less than straightforward. Haythorn (1961) notes the difficulty that operations analysts and operations researchers often encounter in trying to analyze decision situations in complex organizations to the point that prescriptive models can be applied. He ascribes the difficulty to several factors: "First is the fact that organizations are constructed by men with some purposes in mind, although these are not usually stated very explicitly. Analytic solutions must assume that the decision maker is rational, that the parameters relevant to the decision are quantifiable, and that the information necessary to make an optimum decision is available. A careful look at the view of the world held by critical decision makers reveals that they are by no means completely rational; that some of their objectives are not easily quantifiable, and perhaps even incompatible with other objectives; that they do not have all of the information needed in many cases; and that frequently the information they have is inaccurate" (p. 23).

Schrenk (1969) has argued that progress on the development of techniques for aiding decision makers will be impeded until a model of "optimum" decision processes that makes realistic assumptions about human capabilities is forthcoming. Such a model, Schrenk suggests, should reflect the behavior of "reasoning man," a concept that he distinguishes from the rational man of economic decision theory. "The idea is not to specify an 'ideal' decision procedure which will produce perfect choices in abstract or laboratory situations, but rather to develop a process that will yield better decisions in real situations" (p. 548). Schrenk

sees four purposes that such a model might serve: (1) it could provide a framework for the classification and integration of the results of decision-making research; (2) it could provide guidance for further research; (3) it could help system designers to structure decision tasks and to allocate decision functions to men and machines; and (4) it could help guide the development of decision-aiding concepts.

2.2 Worth, Probability and Expectation

Sometimes a decision maker has the task of choosing one from among several alternative courses of action, knowing what the effect of any choice would be. (This situation, which is referred to as decision making under certainty, is discussed in Section IX.) Often, however, one must make a choice when the consequences of that choice cannot be anticipated with certainty. In the latter situation, the decision maker is said to be making a decision "under risk." The most common way of dealing with risky decisions quantitatively has been with models that make use of the concept of mathematical expectation.

The "expectation" associated with a choice is calculated by obtaining the product of some measure of worth of each outcome and a measure of the probability of that outcome, and summing over all outcomes that could result from the choice of interest. It has sometimes been assumed that the decision maker attempts to make a choice that maximizes his "expected" gain. More precisely, it is assumed that the decision maker behaves as though he calculated for each action alternative, the sum of the products of the worths and probabilities of the possible outcomes associated with that alternative, and picked the alternative for which this sum was greatest. The "as though" in the preceding statement is important. No one contends that decision makers, as a rule, really perform the arithmetic necessary to compute expectation; it is only suggested that choices are made as though they were based on such calculations.

Each of the factors in the expectation equation--worth and probability--can be treated as either an objective or a subjective variable. The four possible combinations of objective and subjective indicants of worth combined with objective and subjective measures of probability define four classes of expectation models that have been studied. Table 1 gives expressions, in the notation used by Coombs, Bezembinder, and Goode (1967), for expectations representing each of these models. Much of the research on decision making under risk has been concerned with determining which of these models is most descriptive of human behavior, and with developing techniques for measuring subjective worths and probabilities.

TABLE 1. FOUR BASIC TYPES OF EXPECTATION MODELS.

Model	Type of worth measure	Type of probability measure	Expectation associated with j^{th} possible outcome	Expectation associated with k^{th} choice (which has n possible outcomes)
Expected value	objective	objective	$p_j \ v_j$	$\sum_{j=1}^{n} p_j \ v_j$
Expected utility	subjective	objective	$p_j \ u_j$	$\sum_{j=1}^{n} p_j \ u_j$
Subjectively expected value	objective	subjective	$\psi_j \ v_j$	$\sum_{j=1}^{n} \psi_j \ v_j$
Subjectively expected utility	subjective	subjective	$\psi_j \ u_j$	$\sum_{j=1}^{n} \psi_j \ u_j$

---

p_j : an objective probability

ψ_j : a subjective probability

v_j : an objective measure of value (e.g. amount of money)

u_j : a subjective measure of worth (or utility)

The first of the models listed in Table 1, the Expected Value model, is the least complex conceptually, and the most easily applied, inasmuch as both of its parameters are objectively defined. Although this model has some appeal as a prescriptive model, it has proved not to be generally descriptive of how real decision makers behave (see, for example, Coombs, Dawes, & Tversky, 1970; Edwards, 1961; Lichtenstein & Slovic, 1971; Lichtenstein, Slovic, & Zink, 1969).

The inadequacy of the Expected Value model as a descriptive model is clearly illustrated by the well-known St. Petersburg paradox. Suppose one were offered an opportunity to purchase the following gamble. A fair coin is to be tossed until it comes up tails, at which time the coin tossing is terminated and the winnings are collected. If the coin comes up heads on the first toss, the purchaser will receive \$2.00; if it comes up heads on both the first and second toss, he will receive \$6.00 (or \$2.00 for the first toss and \$4.00 for the second). More generally, if it comes up heads for n consecutive tosses, he will receive \$2.00 for the first toss, \$4.00 for the second, \$8.00 for the third,... and $\$2^k$ for the kth, for a total of

$$\sum_{k=1}^{n} 2^k \text{ dollars.}$$

Since, by definition, the successive tosses are independent, the expected value of this gamble in dollars is given by

$$EV = \frac{1}{2} \bullet 2 + \frac{1}{4} \bullet 4 + \ldots + \frac{1}{2^n} 2^n + \ldots = 1 + 1 + 1 + \ldots$$

which is to say, it is infinite. If one were attempting to maximize expected value, therefore, one should be willing to pay a large amount of money indeed to play this game. It would be surprising, however, if many people could be found who would be willing to risk their life savings, say, which would be small by comparison with the expected gain, to purchase this gamble. In general, it is clear that the attractiveness of a gamble depends not only on the expected value of the outcome but on such factors as the amount that one could possibly lose, and the nature of the distribution of probabilities over the possible outcomes. In the gamble described above, for example, the probability is .5 that the purchaser will win nothing, and .75 that he will win at most \$2.00.

In spite of the inadequacy of the Expected Value model as a generally valid description of behavior, it should be noted that the model does a creditably good job of describing behavior, at least grossly, in many decision situations. Even in the case of gambling behavior, it does not invariably fail; "about 88% of the job" of explaining the behavior of the Las Vegas gamblers studied by Edwards, for example, could be done on the basis of a knowledge of the expected value of each bet (Rapoport & Wallsten, 1972).

Implicit to the Expected Value model is the assumption that the monetary value of a decision outcome represents its real worth to the decision maker, and that this worth is the same for all individuals. Recognition that such an assumption is undoubtedly false led to the formulation of the Expected Utility model in which monetary value is replaced by a measure of the "utility" of an outcome for the particular decision maker involved. According to this formulation the same decision outcome may appeal to different individuals to different degrees, and, consequently, preferences among decision alternatives with uncertain outcomes may differ from one decision maker to another. The Expected Utility model was first proposed by Bernoulli (1738) and given its modern axiomatic form by von Neumann and Morgenstern (1947).

Given that the worth factor in the expectation equation is defined as a subjective variable, the question arises concerning how probability should be defined. Although a review of the controversy would take us too far afield, it should be noted that the question of what the concept of probability "really means" has been the subject of endless philosophical debate. It is sufficient for our purposes to recognize that statements of the type "the probability of the occurrence of event X is equal to Y" have been used in a variety of ways. Such a statement is sometimes used to refer to the relative frequency with which X has been observed over the course of many similar situations. Or it can have reference to a ratio in which the numerator represents the total number of ways in which the outcome of an hypothetical experiment can satisfy some criterion and the denominator represents the total number of different outcomes (as when one says the probability of rolling a 2 or less on a fair die is 2/6).* Sometimes a probability statement is used to refer to the strength of one's confidence, or the degree of one's belief, that an event X, as opposed to the other events that are considered possibilities, will occur. It is this connotation that we here refer to as "subjective probability."

In some situations it makes little if any practical difference which of these connotations one gives to the concept of probability,

*Related to this usage of the term is the so-called "Principle of Insufficient Reason," which directs the decision maker to consider all possible outcomes to be of equal likelihood in the absence of information which indicates such a consideration to be inappropriate. See Rapoport (1964) for an interesting discussion of the limitations of this prescription in defense of an assertion that the six faces of a die are equally likely when one has no reason to assert otherwise.

inasmuch as they will all yield the same numbers. Most people would perhaps agree, for example, that the probability of tossing heads on a fair coin is .5, irrespective of their philosophical position concerning how probability should be defined. Many "probabilistic" situations of interest to investigators of decision making do not easily admit of an analysis in terms of relative frequencies, or even of theoretical ratios, however, and it is perhaps for this reason that many decision theorists subscribe to the notion that probability is best defined in terms of degree of belief. Rapoport (1964) defends this position the following way. "We a told that decisions involving the probability of the outbreak of a nuclear war are based on 'calculated risks,' by which term those who recommend or make decisions must imply calculations involving probabilities. Since the probability of an event such as the outbreak of a nuclear war can have nothing to do with the frequency of such events (since at this writing none has occurred, and, in all likelihood, no more than very few can occur), either the phrase 'the probability of a nuclear war' has no meaning at all, in which case the notion of the 'calculated risk' is only eyewash, or else 'probability' has another meaning, having nothing whatsoever to do with frequency" (p. 25).

The argument that probability often cannot be defined meaningfully in terms of relative frequencies or ratios is a strong one for resorting to a definition in terms of subjective uncertainty. Even when an objective definition is easy to come by, however, one may question whether it should be used by any theory that purports to be descriptive of the behavior of real decision makers. It is the decision maker's own expectation that is presumably important in determining his behavior and his expectation must be calculated in terms of the probabilities as he perceives them. Moreover, it is required of a rational man that his behavior be consistent with the information at his disposal, but not that he have perfectly accurate information. Thus, two decision makers could behave optimally, but quite differently, in the same situation if their perceptions of the situation differed, a fact that is easy to accommodate when probability is defined as degree of belief but not when it is defined strictly in terms of the objective details of the situation.

In the foregoing discussion of Expected Value and Expected Utility models, it was tacitly assumed that the probability factor in the expectation equation was objectively defined. As suggested by Table 1, two additional types of expectation models might be realized by combining subjective probabilities with both objective and subjective measures of worth. The resulting models might be referred to, respectively, as Subjectively Expected Value and Subjectively Expected Utility models. Although both of these types of models have been considered, the latter is by far the more widely accepted and used. This model has been presented by Savage (1954) and by Edwards (1955). Among the four models listed in Table 1 which

have been referred to as single-stage algebraic decision models--it has received the greatest amount of empirical support, and at the moment, ranks as the most influential (Rapoport & Wallsten, 1972).

Savage's (1954) formulation of decision theory identifies a number of "seemingly agreeable" (Tversky, 1969) rules that should be satisfied before it is appropriate to assign a single fixed number denoting worth to each possible decision outcome and a single fixed number denoting judged likelihood of occurrence and then to select maximum products. These rules (see Becker & McClintock, 1967) are as follows:

Rule 1: Transitivity. If, in a choice situation, the decision maker prefers Outcome A to Outcome B and Outcome B to Outcome C, he should prefer Outcome A to Outcome C.

Rule 2: Comparability. The decision maker should be willing to compare two possible outcomes and decide either that he prefers one to the other or that he has no preference between them.

Rule 3: Dominance. If the decision maker determines that, under every possible condition a choice of one of his alternative actions results in an outcome at least as desirable as that which would result from the choice of a second alternative action, and results in a more desirable outcome under at least one possible condition than would the second action, the second action should not be preferred to the first.

Rule 4: Irrelevance of nonaffected outcomes. If the decision maker determines that, for a particular state of the world, two or more of the actions open to him result in the same outcome, his preferences among such actions should not be affected by the outcome associated with that state.

Rule 5: Independence of beliefs and rewards. The decision maker's statement concerning the likelihood of occurrence of a given outcome should not be affected by what he hopes will occur.

Some of these rules seem to be honored as much in the breach as in the observance (see, for example, MacCrimmon, 1968). Violations of Rule 1 are of major significance. This is so because the assumption of transitivity of preferences is a necessary requirement for the construction of a consistent ordinal utility function. (For a discussion of the problem of generating utility functions from preference judgments, see Roberts, 1970.) Tversky (1969) refers to transitivity as "the cornerstone" of decision theory and points out that it underlies measurement models of sensation and value as well. He also notes that decision makers often do violate the transitivity rule in specific situations.

Another rule which seems difficult to satisfy is that requiring independence of beliefs and rewards (Rule 5). MacCrimmon (1968) has found a strong dependency between an individual's estimates of the likelihoods of events and his "tastes"--the worths he assigns to those events. As might be anticipated, such an association may pose difficult analytic problems, since, for a given set of choices, one cannot assume that a distribution of (stated) preferences arises simply out of differences associated with but one of the two parameters in the expectation equation. In principle, this problem is similar to the so-called "conjoint measurement problem" which has received major attention in the context of Subjectively Expected Utility theory.

A variant of the dominance principle (Rule 3) has been stated by MacCrimmon so as to apply to the problem of comparing alternatives that differ with respect to several attributes when preferences can be stated with respect to single attributes individually: "When comparing all alternatives, if some alternative has higher attribute values for all attributes, we say that this alternative 'dominates' the others. We can weaken this notion somewhat and say that if one alternative is at least as good as the other alternatives on all attributes, and is actually better on at least one of them, then this can still be considered the dominant alternative. Conversely, if one alternative is worse than some other alternative for at least one attribute, and is no better than equivalent for all other attributes, then we can say the former alternative is dominated by the latter" (p. 18). Some writers have noted that the dominance criterion is inconsistent with the maximin criterion of game theory (Marschak, 1950; Luce & Raiffa, 1957). Ellsberg (1961) has discussed additional problems with this rule.

Some other assumptions that have usually been considered necessary to the use of expectation models are the following: (1) that the act of gambling has no utility itself; (2) that the subjective probabilities associated with the alternative decision outcomes sum to unity; (3) that preferences are independent of the method by which they are measured. It has not been possible to demonstrate that the first two of these assumptions are simultaneously valid. Moreover, Slovic (1966) and others (Lichtenstein & Slovic, 1971; Lindman, 1970) have shown that preferences among gambles may indeed depend in part on the method by which they are obtained (e.g., a rating procedure versus a bidding procedure). In spite of these limitations, expectation models, and in particular the Subjectively Expected Utility model, have proven to be reasonably predictive of at least certain types of choice behavior (Coombs, Bezembinder, & Goode, 1967). They clearly do not, however, tell the whole story of how to account for human choice behavior.

The demonstration that expectation models such as those described are unable to account for choice behavior consistently and completely has led some theorists to seek to modify (which has invariably meant to complicate) the models to make them more descriptive. Other theorists have simply rejected them out of hand. Payne (1973) points out that models such as those we have considered involve the representation of risky alternatives as probability distributions over sets of decision outcomes, and attribute the choice among the decision alternatives to some function of each distribution's central tendency. In the hope of developing models with greater predictive power, some theorists have looked not only to central tendency measures, such as expected or mean values, but to variances and higher moments of these distributions as well (Becker & McClintock, 1967). Still others have made modifications that relax the requirement that the decision maker's choice be invariably dictat d by which of his alternatives represents the greatest expectation; "random utility" models have been proposed, for example, which assume that the utility of a given outcome is a random variable and that variations in this variable produce variations in choice (Becker, DeGroot, & Marschak, 1963).

Shackle's (1967) assessment of expectation models is representative of the opinions of theorists who reject such models out of hand. He argues that the concept of mathematical expectation, and, indeed, the concept of probability as well, are irrelevant to the assessment of one-of-a-kind decision situations. Furthermore, he contends, most real-life decision situations of interest are, to those who face them, unique events; never before has the individual been called upon to make exactly the choice that he faces and never again will he have to select from among the same set of action alternatives under precisely the same circumstances. In such cases, Shackle argues, the decision maker is concerned with what <u>can</u> happen as a result of his choice, not with what <u>would</u> happen if the experiment were repeated a large number of times: "he is concerned with <u>possibility</u> and not <u>probability</u>" (p. 40). We should note that the argument implies a relative-frequency connotation of probability, a connotation that not all decision theorists accept.

Miller and Starr (1969) suggest that one can always find a way to view a decision problem as a maximization problem if one wants to do so: the quantity that the decision maker wishes to maximize is the degree of attainment of his objective. But this is not very helpful as a definition: indeed, it comes close to being tautological. Miller and Starr apparently do not intend to assert as an empirical fact that decision makers do attempt to maximize anything. More generally, whether decision makers attempt to find <u>optimum</u> solutions to their decision problems Miller and Starr consider to be questionable. Simon (1955) has taken the position that they usually do not. According to his "principle

of bounded rationality" what they do instead is to define a limited set of acceptable, or "good enough," decision outcomes and then select a strategy that they consider to be likely to achieve one of these.

The current status of expectation models among investigators of decision making is reasonably well summarized by three observations. (1) The models that are seriously advocated as descriptive of human behavior are rather more complex than the straightforward Expected Value model that was originally proposed. The history of the development of expectation models may be fairly characterized as a progression from the simple to the more complex: objectively defined variables have been replaced with variables defined in subjective terms, and the number of model parameters has been increased. (2) Even the most complicated models have not proven to be totally descriptive of behavior and some theorists have challenged the validity of the basic assumption of this class of models, namely that the decision maker is motivated to maximize an expectation, no matter how the factors from which expectation is computed are defined. (3) Their limitations notwithstanding, expectation models--even the least sophisticated Expected Value model--do a reasonably good job of predicting choice behavior in many situations. The challenge is to come up with models that can handle the situations for which these models fail, as well as those for which they succeed. Meanwhile, when the maximization of expectation is recognized as the decision objective, then expectation models can be used prescriptively to guide the decision process.

2.3 Game Theory

The theory of games was developed to deal with situations in which the outcomes of an individual's decisions depend not only upon his own actions but also upon those of one or more "opponents" --decision makers whose objectives conflict to some degree with his own. Of special interest is the so-called "zero-sum" situation in which the worths of the outcomes to the opponents sum to zero; one loses what another wins. A commonly prescribed strategy for each "player" of a zero-sum game is to make choices in such a way as to minimize his maximum possible loss, the so-called minimax rule.

The assumptions of game theory are open to a number of criticisms. Shackle (1967), for example, characterizes the theory of games, as developed by vonNeumann and Morgenstern, as "essentially a study of the logic of how to present as impregnable a front as possible to an infallibly wise and rational opponent" (p. 61). The assumption that one is in a conflict and that one's opponent is rational and infallibly wise leads directly to the minimax doctrine. Shackle questions to what extent this conceptualization

can be taken as a reasonable approximation to reality. "Is the impersonal world of nature or even that of business actively concerned to defeat us? Is the human opponent reasonably assumed to be infallible? Is there no essential and ineradicable uncertainty in the outcomes of such few big experiments, large in time scale in comparison with the human life-span, that any of us has time to make? Rather than minimax our losses, is it not more reasonable to fix for them some maximum tolerable numerical size, to avoid any action-scheme which would bring losses larger than this within the range of possible or 'too-possible' outcomes, and subject to this constraint to choose that action-scheme which brings within the range of possible or 'sufficiently possible' outcomes, as high a positive success as we can find?" (p. 65).

In a similar vein, Becker and McClintock (1967) question what they refer to as game theory's "principle psychological assumptions." They point out that the theory assumes, on the one hand, that both decision makers will attempt to maximize their own utility and, on the other hand, will attempt to minimize their maximum losses. These assumptions are inconsistent unless the decision makers look at the game from each other's points of view--a requirement which Morin (1960) finds unsupportable on empirical grounds--and unless the utilities of each decision maker are known to the other and sum to zero for each possible outcome.

Despite its limitations, game theory has provided a valuable framework within which to view decision making in such fields as economics, political science, social psychology and military strategy. The theory has been extended to cover non-zero-sum situations, situations permitting cooperation or collaboration among subsets of players of multiperson games. In addition to minimax, other strategies have been identified as either prescriptively appropriate, or descriptive of behavior, in particular situations.

A short and very readable exposition of the basic concepts of game theory may be found in Edwards (1954). A comprehensive tutorial treatment is provided by Luce and Raiffa (1957).

2.4 Decision Theory and Training

It is a reasonable question to raise whether one may hope to be an effective decision maker in a variety of situations without some intellectual appreciation for the decision-making process, as it is represented by theoretical treatments of decision making. One would guess that there would be some advantage to being familiar, at least with certain of the key concepts that decision theorists employ. In practice, this would mean providing would-be decision makers with a basic introduction to probability theory as well as a working familiarity with notions of rationality, value, utility, mathematical expectation, risk, risk preferences, and so on.

In fact, one could make the case that failure to provide an adequate grounding in theory might deprive the decision maker of the sorts of insights that would lead to productive use of available decision-aiding techniques. The demonstration by MacCrimmon (1968) that decision aids developed in quite disparate contexts can be effectively brought together in the solution of problems involving multi-attribute alternatives, suggests the utility of broad acquaintance with basic concepts and principles.

In reporting one effort to develop a system to assist corporate decision makers by enabling them to manipulate parameters (entered as distribution functions) on preprogrammed tree models, Beville, Wagner, and Zanatos (1970) made some observations that are relevant to this point. They noted that the use of subjective probability distributions as inputs to models is novel even to experienced decision makers, and must be carefully taught. More generally, they concluded that a black-box approach to utilization of the system would have been markedly inferior to one in which the workings of the system were explained to the user.

The teaching of decision theory should, of course, distinguish what is intended to be prescriptive from what is considered descriptive of the behavior of human decision makers. It should also clearly identify the limitations of the models that are considered. Tutorial treatments of decision theory and game theory are readily available sources of training material (Edwards, 1954; Edwards & Tversky, 1967; Howard, 1968; Lee, 1971; Luce & Raiffa, 1957; Miller & Starr, 1967; North, 1968; Rapoport, 1960; Schlaifer, 1969). A comprehensive bibliography of research reports has been prepared by Edwards (1969).

Whether familiarization with theoretical treatments of decision making will in fact improve decision-making behavior is a question for empirical research. Our guess is that the answer will be a qualified yes. Such training will be efficacious for some people performing certain types of decision tasks but perhaps not for all people or all tasks. One objective of training research should be to identify those conditions under which such training would be effective and those under which it would be a waste of time.

SECTION III

CONCEPTUALIZATIONS OF DECISION SITUATIONS AND TASKS

Numerous ways of conceptualizing decision processes have been proposed by different investigators. Some conceptualizations emphasize differences among decision situations; others focus on the tasks that decision makers are required to perform. All of them have the same purpose, namely that of simplifying the problem of thinking about decision making by identifying a few "types," each of which is representative, in terms of some critical aspects, of a variety of specific situations or tasks. We review briefly in this section a number of proposed simplifying conceptualizations. There is no attempt to be exhaustive. The intent is simply to illustrate by means of a few examples some of the ways in which investigators of decision making have characterized or categorized the object of their study.

3.1 Classifications of Decision Situations or Decision Types

3.1.1 Edwards

Edwards (1967) makes a distinction between static and dynamic decision situations. In the former case, a one-time decision is required, whereas in the latter, sequences of decisions are made, earlier decisions and their outcomes having implications for subsequent ones. Six types of dynamic decision situations are distinguished on the basis of such factors as whether the environment is stationary or nonstationary, whether or not the environment is affected by the decisions that are made, and whether or not the information about the environment is affected or controlled by those decisions. Edwards further classifies psychological research relating to decision making under four topics: information seeking, intuitive statistics, sequential prediction, and Bayesian processing.

3.1.2 Howard

Howard (1968) characterizes decision situations in terms of three orthogonal dimensions: degree of uncertainty, degree of complexity (number of relevant variables), and degree of time dependence. The various combinations of the extreme values on these dimensions are taken as representative of eight prototypical situations, for each of which there is an appropriate set of analytical tools. An example of a deterministic (no uncertainty), single variable, static (time-independent) problem would be to determine the largest rectangular area that can be enclosed with a fixed amount of fencing. The appropriate mathematical tool would be the calculus. Decision problems like assigning customers to warehouses or jobs to men would, in Howard's taxonomy, be in the category defined as deterministic, complex (many variables), and

static. Matrix algebra and linear optimization are appropriate mathematical techniques.

3.1.3 Sidorsky

Sidorsky and his colleagues have proposed a taxonomy of types of decisions encountered in tactical military situations (Sidorsky, Houseman, & Ferguson, 1964; Sidorsky & Simoneau, 1970; Hammell & Mara, 1970). The acronym ACADIA is used as a mnemonic for the six types of "situational demands" identified by the taxonomy: Acceptance, Change, Anticipation, Designation, Implementation, and Adaptation.

An acceptance-type decision has to do with applying data to the acceptance or rejection of a hypothesis concerning some characteristic of the enemy. Detection, classification and localization are associated operations or objectives. The acceptance-decision idea seems to be close to what some other investigators have referred to as situation diagnosis. A change-type decision involves the decision maker in a choice between initiating a new tactical operation or continuing the course of action on which he is already launched. An anticipation-type decision is required when a decision maker must predict what the state or intention of an enemy force will be sometime in the future.

A designation-type decision involves the choice of one from among a set of possible action alternatives. An implementation-type decision has to do, not with the selection of an action alternative, but with the determination of the proper time to execute it. An adaptation-type decision is called for when the decision maker is faced suddenly with unexpected and perhaps potentially disastrous circumstances.

3.2 Classifications of Decision Tasks

3.2.1 Howard

Howard conceives of the decision process as being composed of three phases: (1) the deterministic phase, (2) the probabilistic phase, and (3) the information phase. In the deterministic phase, the decision analyst identifies the state and decision variables and constructs a model of the decision problem. In the probabilistic phase, he assigns probability distributions on the state variables. In the information phase, he determines what additional information should be gathered to reduce uncertainty further. Howard estimates that the first phase represents about 60% of the total effort of the decision maker, while the second and third phases represent about 25% and 15%, respectively.

3.2.2 Adelson

A taxonomy of decision tasks that are carried out in modern military command-and-control systems is proposed by Adelson (1961). Four types of tasks are distinguished: (1) characterization of the state of the world, (2) determination of the available action alternatives, (3) outcome prediction and (4) choice rationalization. The first task type refers to the need of the decision maker to characterize the current state of the world in a way that is relevant to his decision problem. The definition of the variables in terms of which the characterization should be made, and the assessment of the relative stability of the world that is being observed are seen as significant problems. The second task type acknowledges the need to make explicit the courses of action that are open to the decision maker. The difficulty of this task may depend somewhat on how rapidly the situation is changing and on the cost of obtaining information. Outcome prediction refers to the process of attempting to anticipate what the consequences would be if specific action alternatives were selected. The final task type involves the need to justify one's choice of action in terms of the objectives of the command-and-control system.

3.2.3 Drucker

Drucker (1967) has identified six steps that he considers to be involved in the process of making the types of decisions that confront business executives: (1) the classification of the problem, (2) the definition of the problem, (3) the specifications which the answer to the problem must satisfy, (4) the decision as to what is "right (as distinguished from what is acceptable in order to meet the boundary conditions), (5) the building into the decision of the action to carry it out, and (6) the feedback which tests the validity and effectiveness of the decision against the actual course of events.

3.2.4 Soelberg

Soelberg's (1966) taxonomy, like Drucker's identifies six aspects of the decision making process: (1) problem recognition, (2) problem definition, (3) planning, (4) search, (5), confirmation and (6) implementation.

3.2.5 Hill and Martin

A model proposed by Hill and Martin (1971) also recognizes six different categories of activities in the decision-making process: (1) identification of concern, (2) diagnosis of situation, (3) formulation of action alternatives, (4) test of feasibility of selected alternatives, (5) adoption of alternative, and (6) assessment of consequences of adopted alternative. The model assumes that

the decision maker's behavior at each of these steps is influenced by what he knows of the theory and practice of decision making as well as by what he knows about the setting in which the decision problem exists. Hill and Martin identify nineteen skills that they consider to be implicit in these six generic activity categories:

"1. Asking for and receiving feedback
2. Assembling the facts (including past experience as it bears on the decision)
3. Identifying the courses of action available
4. Identifying forces for and against the alternatives
5. Ranking and rating alternatives (includes putting a value on applicable risk factors)
6. Assessing the people-task ratio
7. Identifying the latest and expected consequences of the alternative courses of action
8. Determining the advantages and disadvantages of each action alternative
9. Testing the validity and effectiveness of the consequences of the decision against the actual course of events to evaluate the decision maker's judgment and to modify his subsequent decision-making behavior
10. Brainstorming action alternatives
11. Classifying and defining the problem requiring a decision
12. Analyzing and evaluating stimuli and decisions coming in from the outside
13. Defining the goal at which the decision is directed
14. Communicating the decision in written or verbal composition
15. Identifying resources bearing on the making of the decision
16. Recognizing the need for a decision
17. Utilizing minor, relatively simple decisions to contribute to making the more complex one (includes determining the hierarchy of order in which minor decisions will be dealt with and coping with timing as alternatives come into focus and seemingly demand attention at the same time)
18. Obtaining information
19. Specifying the boundary conditions the decision must satisfy" (p. 433).

3.2.6 Edwards

Edwards (1965b) lists the following thirteen steps that must be carried out by any Bayesian decision system:

"1. Recognize the existence of a decision problem

2. Identify available acts

3. Identify relevant states that determine payoff for acts

4. Identify the value dimensions to be aggregated into the payoff matrix

5. Judge the value of each outcome on each dimension

6. Aggregate value judgments into a composite payoff matrix

7. Identify information sources relevant to discrimination among states

8. Collect data from information sources

9. Filter data, put into standard format, and display to likelihood estimators

10. Estimate likelihood ratios (or some other quantity indicating the impact of the datum on the hypotheses)

11. Aggregate impact estimates into posterior distributions

12. Decide among acts by using principle of maximizing expected value

13. Implement the decision" (p. 142, Table 1).

Steps 1 through 5, and 7 and 10, Edwards suggests, are best performed by men, Steps 6, 11 and 12 by machines, and Steps 8, 9 and 13 by both men and machines. Steps 1 through 7 may be done in advance of the decision time; Steps 8 through 13 must be done at the time that the decision is to be made. (See Section VIII for a discussion of Bayesian information processing.)

3.2.7 Schrenk

A conceptualization of the decision-process that we find particularly interesting is one proposed by Schrenk (1969). The motivation for developing this conceptualization was to provide a representation of the decision-making process that is prescriptive in the sense that it can be used as a guide for the structuring of decision-making tasks of man-machine systems, but which does not make unrealistic assumptions about human capabilities. The conceptualization is viewed by Schrenk as tentative, and in need of further development; however, even as it stands it provides the system designer with a great deal of food for thought concerning how to allocate decision functions among men and machines.

Three major categories of decision tasks, or phases of the decision process are distinguished: (1) problem recognition, (2) problem diagnosis, and (3) action selection. Each of these phases is further broken down into several components, and flow-diagrams are given which show where the components appear in the overall process. The following is a paraphrasing of Schrenk's description of each of these components.

- Problem Recognition: Determination that a problem requiring a decision exists.
  - Acquire information: Receipt of information indicating that actual situation differs from the desired situation.
  - Recognize objectives: The decision maker's purpose or mission.
  - Perceive decision need: Perception of difference between objectives and current situation; may result from change in situation or in objectives.
  - Assess problem urgency and importance: Establishment of priority of problem, relative to other problems demanding attention, and allocation of resources for solving it.
- Problem Diagnosis: Determination of the situation that is causing problem.
  - Define possible situations: Generation of hypotheses regarding situation.
  - Evaluate situation likelihoods: Assignment of *a priori* probabilities to alternative hypotheses.

- Determine whether more information is needed: Assessment of adequacy of information in hand; a continuing process.
- Identify possible data sources: If more information is desired.
- Judge value versus cost: To determine whether, or how, desired information should be acquired.
- Seek more information: Assuming value judged to be greater than cost.
- Re-evaluate situation likelihoods: Iterate.
- Determine whether alternatives under consideration account for all the data: Recognition of possible need to modify set of hypotheses being considered.
- Make diagnostic decision: Selection of favored hypothesis, or possibly of small set of weighted alternatives.

• Action Selection: Choice of course of action.

- Define action goals: Specification of explicit goals, including interim or subordinate objectives.
- Specify value and time criteria: Identification of relevant dimensions of multidimensional goals and specification of time constraints within which decision must be made.
- Weight decision criteria: Establishment of relative importance of various decision criteria.
- Specify risk philosophy: Specification of strategy of action selection insofar as it is dictated by considerations of balancing risks against potential gains.
- Input operating doctrine: Consideration of any rules or doctrine by which the decision maker's behavior should be guided.
- Generate action alternatives: Explicit listing of reasonable set of courses of action open to decision maker.
- Predict possible outcomes: Specification of the possible outcome associated with each of the potential action alternatives.
- Estimate outcome gains and losses: Determination of value of possible decision outcomes.
- Estimate outcome likelihoods: Estimation of probabilities of occurrence of possible outcomes for each action alternative.

- Evaluate expected values of actions versus their costs: Derivation, from preceding two steps, of expected value of each possible action, and estimation of associated cost.
- Evaluate actions by risk philosophy: Assessment of each action alternative in terms of its implications for the risk philosophy that the decision maker has adopted.
- Determine whether more information is needed: As under Diagnosis; a continuing question; new information might be useful either for identifying additional action possibilities, or to improve predictions concerning possible decision outcomes.
- Seek information: If desired, and worth cost of acquisition.
- Re-evaluate action alternatives: Iterate
- Determine whether best action is acceptable: Review of most desirable action alternative to assure its acceptability, in terms of the decision goals and criteria, the expected gains from the choice and the cost of making it.
- Choose course of action: The "decision."
- Implement action: Initiation of whatever steps are necessary to assure that the selected action is carried out.

The main fault that we have to find with Schrenk's model is that it may be overly elaborate. It is doubtful that many individuals go through anything approaching this multistep procedure in the process of making a decision. This is perhaps an unjustified criticism, inasmuch as Schrenk intended the model to be more prescriptive than descriptive. And whether such a model can serve as a prototype procedure for decision makers to follow remains to be seen. In any case, the representation does serve the useful function of making explicit many of the aspects of decision making and it stands as a reminder that decision making may be viewed as a complex and multifaceted process indeed.

3.3 Decision Making as a Collection of Problem-Solving Tasks

We take the position that decision making is best conceived as a form of problem solving; or, more specifically, that it involves a variety of aspects each of which may be viewed as a problem-solving task in its own right. In the most general terms, the decision maker's problem is to behave in a rational, or at least a reasonable, manner. To be sure, the distinctive characteristic of the specific problems with which the decision maker deals is the element of choice; he must at some point decide upon one from among two or more alternative courses of action. While the

act of choosing among alternatives is central to decision making, it is by no means the only problem--or even necessarily the most difficult one--that the decision maker must solve. We wish to emphasize the importance of making explicit the other things that must be done if one is motivated to make the best possible--or at least a satisfactory--decision, given the resources at one's disposal. In many real-life situations, the problem of choosing among possible courses of action is far simpler than that of discovering what one's options are in the first place, or of assigning preferences to possible decision outcomes in a consistent way. Also, the decision maker may find it necessary to make many preliminary decisions simply by way of setting the stage for making the decision which is his primary concern. For example, he will want to reduce his uncertainty about the decision situation or about the consequences of the various choices that are open to him. However, the acquisition of information takes time, and may be costly in other ways, so he will continually be faced with the problem of deciding whether any additional information that he may wish to get is worth the cost of getting it.

It is clear from the foregoing that there are many ways to classify the various tasks that the decision maker may be required to perform. The scheme that we find most satisfactory recognizes eight aspects of decision making: information gathering, data evaluation, hypothesis generation, problem structuring, hypothesis evaluation, preference specification, action selection, and decision evaluation.

This conceptualization has an element of arbitrariness about it--as does any other. There are four points that we would like to make in this regard. First, the decision to conceptualize the process in terms of eight types of tasks, as opposed to some other number, is itself somewhat arbitrary, and reflects our own biases concerning what constitutes a useful level of organization. One might conceptualize the decision process at a much coarser level and distinguish two major types of tasks--diagnosis and action selection--that would encompass all of those that we wish to distinguish. This approach has been taken by several investigators (Bowen, Nickerson, Spooner & Triggs, 1970), Kanarick, 1969; Williams & Hopkins, 1958). Bowen et al. (1970) point out that in the military, diagnosis is the proper function of intelligence, and

action selection that of command. At the other extreme, one might attempt a much finer grained representation and identify a much larger number of activities that a decision maker may be called upon to perform. In this case, each of the tasks we have identified might be replaced with several more detailed tasks. These are not mutually exclusive approaches, of course, and we will have occasion to consider how some of the tasks we have identified may be further broken down. However, this level of analysis appears to us to be the most useful one for our present purpose, and possibly for serving as a general framework in terms of which to think about decision making as a whole.

Second, our taxonomy is not orthogonal to other conceptualizations such as those discussed in the preceding section. It has elements in common with most of them. Indeed, the intent is not to take issue with other taxonomies, but to propose one that represents what, in our view, are the best aspects of all of them.

Third, we do not mean to suggest that whenever an individual finds himself performing the role of a decision maker he explicitly runs through this set of tasks in serial fashion, or even that he performs each of these tasks explicitly at all. Moreover, when he does perform these tasks it is not necessarily the case that he is fully aware of doing so. It is characteristic of human beings that they often can solve problems quite effectively without having any clear idea how they do it. This characteristic has been a frustration to researchers in artificial intelligence, who have found it exceedingly difficult to program computers to perform some tasks that human beings seem to be able to perform with ease. What we do mean to suggest by the proposed taxonomy is that all of these types of activities are implicated in decision making and that any attempt at a thorough discussion of the decision-making process must take account of them.

Finally, viewing decision making as a problem-solving process that is composed of several phases or subprocesses emphasizes the fact that in any given decision situation, different decision tasks could be performed by different individuals or groups (or machines). An implication for training is that it may be less appropriate to think of training decision makers per se than of training individuals to play specific roles in the decision-making process. On the other hand, there will undoubtedly always be some situations in which all the various aspects of a decision problem will be handled by the same individual. But whatever the case, there is perhaps somethi g to be gained by making decision makers--or specialist members of decision-making groups--aware of the many facets of the general task.

In the next few sections of this report, we consider each of the components of our task taxonomy in turn. The order in which the tasks are discussed represents a natural progression; however, in real life decision situations, an individual, in a decision-making system, may perform several of these tasks more or less simultaneously. Or he may skip from one to another in a variety of orders, and may perform any given type of task many times in the course of attempting to solve a single decision problem.

SECTION IV

INFORMATION GATHERING

From the point of view of the decision maker, most decision situations are characterized by some degree of uncertainty. This uncertainty may involve the current "state of the world," the decision alternatives that are available, the possible consequences of selecting any given one of them, and even the decision maker's preferences with respect to the possible decision outcomes. One of the major problems facing the decision maker, therefore, is that of acquiring information in order to reduce his uncertainty concerning such factors, thereby increasing his chances of making a decision that will have a desirable outcome.

What makes the problem interesting, and nontrivial, is the fact that information acquisition can be costly, both in terms of time and money. Therefore, the decision maker must determine whether the value of the information that could be obtained through any given data-collection effort is likely to be greater than the cost of obtaining it. And therein lies a decision problem in its own right.

In theory, one can see an infinite regress here. In order to decide whether to initiate any information-collecting effort, one must determine the worth of the information to be collected and the cost of collecting it. But in order to determine that, one may have to collect some information--at some cost, and so on. In practice, of course, infinite regresses never occur; and in this case, one very quickly gets to a point at which the decision maker relies on information in hand, or appeals to his own intuitions.

4.1 Information Seeking versus Information Purchasing

Information gathering may be thought of as involving two quite different activities: (1) information seeking (locating the information that one needs or wants), and (2) information purchasing (deciding whether information, the location of which is known, is worth what it will cost to acquire it). This distinction is something of an oversimplification, inasmuch as the act of seeking itself typically involves some cost, and one often must decide whether to incur that cost without any assurance that the search will yield the information that is desired. The aspect of "seeking" that we wish to emphasize, however, is the need for identifying and actively searching out information sources, of finding out where the desired information is and going after it. The term "purchasing" is used to connote a more passive role on the part of the decision maker, the opportunity to acquire information is presented to him and he need only indicate whether or not he wants to avail himself--at some cost--of the information that is offered.

The distinction between information seeking and information purchasing is a useful one because it highlights the fact that experimental studies have focused almost exclusively on the latter process; although investigators often have not made the distinction and have frequently discussed their results as though they had to do with the former. Typically, the decision maker is presented with all the information that he needs--although he may have to decide how much of it to purchase--and the process of seeking information is not studied. The world outside the laboratory is not nearly so accommodating, however, and one must either seek out the information one wants, or go without it. Moreover, studies of information purchasing, while they tell us something about how effectively people can judge the worth of information that is made available to them, shed little light on information-seeking behavior.

Perhaps the main reason why information-seeking behavior has not been widely studied is the difficulty of manufacturing situations in the laboratory that are representative of those faced by decision makers in the real world. In any case, whatever the reasons, information seeking per se has not received the attention from investigators of decision making that it deserves. The experiments that we have reviewed that purport to deal with this topic invariably have actually studied information purchasing as we have defined that term.

4.2 Optional-Stopping Experiments

An experimental paradigm that has often been used to study information-purchasing behavior is one in which the decision maker is provided with the opportunity on each trial either of purchasing more data that are relevant to the decision that he is required to make, or of making the decision. The terms "deferred decision" "optional stopping" and "optimal stopping" have all been used to refer to this paradigm. "Deferred decision" and "optional stopping" connote the fact that the subject in such an experiment has the option on each trial of making a terminal decision or deferring it in order to obtain more data. "Optimal stopping" refers to the fact that when the situation is sufficiently well-structured so that the costs and payoffs associated with possible decision outcomes, the cost and informativeness of data, and the decision maker's objectives are all known, the point can be determined at which information purchasing should be stopped and the decision made. The "optional-stopping" paradigm is to be contrasted both with the more familiar paradigm in which the experimenter determines how much information the decision maker will be given, and what is usually called the "fixed-stopping" paradigm in which the decision maker specifies how much information he wishes to purchase, in advance of receiving any.

Often, in optional-stopping experiments, the required decision concerns the parameters of the distribution from which the observational data are being drawn. For example, one may have to decide whether a sequence of red and black poker chips that one observes is drawn from a population in which the proportion of reds to blacks is, say, 60-40 or 30-70. The question of interest in such experiments is whether the subject's information-purchasing behavior deviates from optimality, and if so, in what way?

What constitutes optimal performance has been worked out for a variety of specific situations (Birdsall & Roberts, 1965; Blackwell & Girshick, 1954; Raiffa & Schlaifer, 1961). For our purposes it suffices to recognize that, in general, the amount of information (number of observations) that should be purchased will vary directly with the magnitude of the costs and values associated with the decision outcomes, and inversely, with the cost and "diagnosticity" of the data that are purchased. Diagnosticity refers to the degree to which the data should reduce the decision maker's uncertainty about which of the terminal decision alternatives should be selected. The diagnostic value of a datum depends on several factors (some of which are discussed in Section VIII), and typically decreases as the number of data that have already been collected increases. A factor that usually is not taken into consideration in optional-stopping experiments but can be critical in real-life situations is the importance of time itself. In some situations the potential consequences of a decision are highly time-dependent. This fact can be incorporated in an optimal-stopping rule by making the cost of an observation, or the stopping criterion, a function of time.

Typically, performance in optional-stopping experiments has been found not to be optimal. Moreover, as illustrated by a study by Green, Halbert, and Minas (1964), the deviation from optimality may be in either direction. In one experiment, Green, et al. found that the number of observations purchased increased with the a priori uncertainty concerning the correct decision--as would be expected of an efficient Bayesian processor--however, subjects tended to purchase too many observations when the a priori uncertainty was maximized by providing no prior information concerning the likelihoods of the correctness of the possible decisions. In combination, the results of these experiments suggest that decision makers may sometimes purchase too much information, and sometimes too little. In particular, it would appear that they may purchase too much information if the a priori uncertainty is small, and too little if the a priori uncertainty is large.

Many investigators have used the optional-stopping paradigm (Becker, 1958; Edwards, 1967; Edwards & Slovic, 1965; Fried & Peterson, 1969; Howell, 1966; Irwin & Smith, 1957; Pitz, 1968, 1969;

Pruitt, 1961; Schrenk, 1964; Snapper & Peterson, 1971; Swets & Birdsall, 1967). The results of most of these studies suggest that although information seeking may approach optimal levels (Becker, 1958; Howell, 1966; Pruitt, 1961), there are reasonably systematic departures from perfect performance. The general finding seems to be that too little information is sought when (theoretically) much is required, and that too much is sought when little is required. The latter finding fits well with the conservatism or inertia effect often noted in studies of Bayesian inference, but the former clearly does not.

A few descriptive models of optional-stopping behavior have been developed (see, for examples, Edwards, 1965a; Pitz, 1968; Pitz, Reinhold, & Geller, 1969). These models have been developed in a Bayesian context (Rapoport & Wallsten, 1972) and tend to be situation specific (see, for example, the "World Series Model" of Pitz, Reinhold, & Geller, 1969).

Noting that most optimal-stopping experiments had been concerned only with the question of when to stop acquiring information from a single source, Kanarick, Huntington, and Petersen (1969) suggested that a more valid simulation of some decision-making situations, e.g., tactical situations, would recognize that the decision maker must deal with information from more than one source. In keeping with this observation, Kanarick et al. did an optional-stopping study in which the decision maker had the option on each trial of acquiring data from his choice of three sources, or of making a terminal decision. The terminal decision that was required involved the presence or absence of an enemy submarine in the vicinity. The information sources differed, both with respect to the cost of obtaining information from them and with respect to the reliability of the information obtained. (The topic of reliability of information will be discussed more fully in Sections V and VIII.) Costs associated with incorrect decisions were also manipulated. Although the behavior of the subjects was consistent with the rational model in many ways--they were willing to pay more for more reliable information; how much information they collected before making a particular decision depended on how bad the consequences would be if that decision proved to be incorrect--performance was less than optimal in several respects. The subjects tended, for example, to consult the most reliable (and most costly) sources less frequently and the less reliable (and less costly) sources more frequently than they should have. Kanarick et al. characterized this behavior as a form of conservatism, "a reluctance to expend the resources necessary to obtain the best information in a choice situation" (p. 382). The subjects also tended to purchase less data in general than they should have, and, consequently, made more incorrect decisions and won fewer points than did a Bayesian model that was used to represent optimal behavior.

Levine and Samet (1973) have also studied information gathering--information purchasing in our terms--in a simulated tactical situation. The scenario was a military action and the subjects' task was to decide which of eight locations was the target of a hypothetical enemy advance. On each trial, a subject could either make a terminal decision or request additional information from each of three intelligence sources concerning the present whereabouts of the advancing force. A sequence of reports from a given source represented the path that the advancing force had taken over a period of time, according to that source. Among the variables that were manipulated were the reliability of the intelligence sources, the degree of conflict among reports from different sources, and the probability that a request for information would yield an updated report (as opposed to a repetition of the preceding report). Performance was sensitive to each of the variables. In particular, fewer reports were requested and decisions were more often correct when all the sources were reliable, and the quality of performance tended to decline as the percentage of the sources that were unreliable was increased. Increasing the degree to which the sources were in conflict also had the effect of decreasing the number of reports requested. (This counterintuitive result may be due in part to the fact that as conflict increased in this experiment, so did the probability that the correct target was indicated by at least one of the sources on a given trial.) The number of requests for reports decreased as the probability that a given report would yield new information increased; the relationship was such, however, that the amount of information (number of updates) received increased with this variable.

In a subsequent experiment, in which the same decision problem was used, Levine, Samet, and Brahlek (1974) varied the rate at which new reports were given to the subject, whether the reports were delivered automatically or in response to the subject's request, the possibility of revising an initial decision and the payoff scheme. In this case, performance was better for the faster rates of information acquisition, but was not highly sensitive to whether the rate was self- or force-paced. Increasing the opportunity for revising a decision had the effect of decreasing the accuracy of first decisions and the subjects' confidence in them.

4.3 Decision Revision and Effect of Commitment on Information Gathering

The results of a few studies suggest that one's information-gathering behavior may be different after making a decision than before, particularly if the making of the decision involves some sort of public acknowledgment or commitment (Geller & Pitz, 1968; Gibson & Nichol, 1964; Pruitt, 1961; Soelberg, 1967). People may require more information, for example, to change a decision than was required to arrive at a decision in the first place (Gibson & Nichol, 1964; Pruitt, 1961). This observation is in keeping with the results of several studies that suggest that evidence that tends to confirm a favored hypothesis is often given more credence than evidence that tends to disconfirm it (Brody, 1965; Geller & Pitz,

1968; Pitz, Downing, & Reinhold, 1967). And sometimes disconfirming evidence may even be misinterpreted as supportive of a decision that has already been made (Grabitz & Jochen, 1972).

The motivation for acquiring information may change, following a decision, from that of trying to increase the probability of making a good decision to that of justifying or rationalizing a decision that has already been made. Soelberg (1967) has concluded from a study of the job-seeking behavior of graduates of the Sloane School that people frequently make an implicit selection from among the existing opportunities, following which "a great deal of perceptual and interpretational distortion takes place in favor of the choice candidate" (p. 29). In a somewhat similar vein, Morgan and Morton (1944) have asserted that people often accept conclusions that are consistent with their convictions without regard for the validity of the inferences on which those conclusions are based, and that "the only circumstance under which we can be relatively sure that the inferences of a person will be logical is when they lead to a conclusion which he has already accepted" (p. 39). We will return to the question of the logicality of thought in Section 8.3).

One suspects that in real-world situations the information-seeking behavior that follows the making of a decision may often differ considerably from that that precedes it. In particular, one would guess that to the degree that the motive of the information seeker is the rationalization of a decision already made, the process would become highly selective as to the sources consulted.

4.4 Quantity of Information and Quality of Decision

It is quite natural to assume that the more data one has that are relevant to a choice that he must make, the better his choice will be. The assumption, without qualification, is not valid (Ackoff, 1967; Fleming, 1970; Hayes, 1964; Hoepfl & Huber, 1970; Sidorsky & Houseman, 1966). It is possible, indeed easy, to provide an individual with more information than he can assimilate and use--especially if he is operating under some time pressure. The point is illustrated nicely by an experiment by Hayes.

Hayes had naval enlisted men make decisions concerning which of several airplanes to dispatch to investigate a reported submarine sighting in a simulated tactical situation. The available airplanes differed with respect to such characteristics as speed, distance of its base from the target, delay before it could take off, quality of its pilot, quality of its radar, and so on. Each characteristic could take on any of eight (not necessarily numerical) "values," which could be ranked unequivocally from best to worst. The number of available airplanes from which a subject had to choose was varied (4 or 8) as was the number of characteristics (2, 4, 6 or 8) on

which he was to base his choice. The effect of the latter variable is of particular interest. Decision time increased markedly with this variable; however, the decision quality--which was defined objectively in two ways--did not. Hayes hypothesized that, other things equal, one's sensitivity to the way two alternatives differ with respect to individual characteristics decreases as the number of characteristics that must be considered increases. Of particular relevance to this review is the fact that Hayes trained a second set of subjects for several days to see if they would learn to make better decisions with the larger amounts of information. Although the quality of decisions was generally somewhat higher after training than before, the relationship between decision quality and number of characteristics on which a decision was based did not change.

We should not conclude from this study that one should never, under any circumstances, be provided with more than a very few items of information that are relevant to any choice that one may have to make. One might conclude, however: (1) that decision makers should be trained to recognize their limitations for assimilating information, and to avoid attempting to operate beyond them, and (2) that to the extent that the functional relationship between the desirability of the various choice alternatives that are open to the decision maker and the values of the factors that determine it is known, the implication of particular sets of factor values should probably be computed, and not estimated by men. The problem of determining, or discovering, such functional relationships is a nontrivial one. (See Section IX.)

4.5 A Conceptualization of Information Gathering in the Real World

What makes the real-world decision maker's task particularly difficult is the fact that the information that he would like to have typically is distributed among a variety of sources. One way of characterizing these sources is in terms of the two properties: degree of passivity and degree of cooperativeness. According to this conceptualization, a source is either active or passive, and either cooperative or uncooperative.

An actively cooperative source--the preferred type--volunteers information, and seeks ways to get it to the decision maker. In the military context, an intelligence officer would be an actively cooperative source for a commander.

A passively cooperative source is one that would provide information if solicited, but does not volunteer it. A possible reason for not volunteering information in this case is a failure of the source to recognize itself as such. An example, again from a military context, would be friendly inhabitants of an area of operations who have information that would be valuable to a

military commander, but are unaware of the fact. The problem that the decision maker has vis-a-vis passively cooperative sources is to identify and find them.

An actively uncooperative source has information that would be of use to the decision maker, but being motivated to thwart the decision maker's objectives if possible, volunteers information that is misleading. A propagandist is an example of such a source. The decision maker's problem with respect to actively uncooperative sources is to recognize them as such and to assess the information obtained from them accordingly.

A passively uncooperative source is one that withholds information from the decision maker, and further will not provide it if asked. Hostile noncombatants in an area of military operations might fit this description, as might espionage agents. The decision maker's problem with respect to passively uncooperative sources is to persuade them to change their status and become actively cooperative. History, both real and fictitious, is replete with accounts of the unsavory methods that have been employed to this end.

To the extent that laboratory studies of decision making have been concerned with information gathering, they have involved actively cooperative sources almost exclusively. The problem of finding sources that are nonobvious and that of coping with those that are noncooperative have received very little attention from experimenters. In part this is undoubtedly due to the fact that capturing the essence of these aspects of information gathering in laboratory situations is a very difficult thing to do. And the alternative of studying these processes in situ is hardly less difficult. Until such studies are performed, however, our understanding of how decision makers go about gathering - especially seeking - information so as to increase their chances of making effective decisions will remain very incomplete.

4.6 Information Gathering and Training

We stress again that laboratory studies of information gathering have failed to capture the complexity of the problem that often faces the information seeker outside the laboratory. Consequently, very little is known about information seeking behavior as it occurs in the real world. This is unfortunate because information seeking constitutes a particularly critical aspect of many real-life decision problems and so long as this behavior is not well understood, our understanding of decision making will be incomplete.

The implications for training are obvious: training procedures that are based on a solid foundation of factual knowledge about human capabilities and limitations cannot be developed if the foundation does not exist. The need is for research that is designed to answer some of the questions that laboratory experiments heretofore have failed to address effectively. Such questions include the following. How good are people at identifying sources of information that is relevant to their decision problems? How do they go about discovering such sources? How capable are they of assessing the cost of acquiring information that may be difficult to get and the worth of the information that might be obtained? To what extent can useful principles and procedures for information seeking be made explicit and taught? It is probably fair to say that with respect to such questions there is insufficient basis for even an educated guess as to the answer. Clearly there is need for some imaginative research on this aspect of the decision-making process.

Laboratory studies such as those reviewed above do shed some light on information <u>purchasing</u> behavior. In particular they tell us something about human capabilities and limitations in assessing the worth of information in well structured situations. Although it would be risky to generalize many of the conclusions uncritically to nonlaboratory situations, the conclusions nonetheless are suggestive of what should perhaps be done by way of training or training research.

SECTION V

DATA EVALUATION

In the preceding section we used the words data and information more or less synonymously. It will be helpful at this point to make a distinction. The term data is perhaps best used to refer to what one collects, and the term information to connote whatever conclusions or inferences one draws from data. The data and the information extracted therefrom can be identical, but they need not be. For example, if a military commander receives data to the effect that the troop strength of an opposing tactical force is 15,000 men, and he considers the source to be a reliable one, he will undoubtedly accept the data as accurate and conclude that the enemy troop strength is indeed 15,000 men. On the other hand, if he has less than full confidence in the source of this report, he may tentatively conclude that the troop strength is somewhere between 5,000 and 25,000 men, and attempt to get more data from which he can derive a more precise estimate.

The point is that as part of the process of attempting to reduce his uncertainty about his decision situation, the decision maker must evaluate the data that he receives as to their pertinence and trustworthiness. In other words, the first decision that the decision maker must make with respect to any new datum is how seriously he should take it. He may not explicitly do this in all cases, but to fail to do so at least implicitly is tantamount to judging his sources as completely trustworthy and their inputs as equally important.

5.1 The Evaluation versus the Use of Data

There are two questions relating to data quality that deserve attention: (1) how well can people judge and report the quality of the data on which decisions are to be based, and (2) how effectively can they utilize information concerning quality of data when that information is provided for them? The first of these questions concerns what we are referring to as the task of data evaluation, and is discussed in this section. The second has to do with data utilization and is more appropriately discussed in connection with hypothesis evaluation in Section VIII.

In anticipation of the latter discussion, we note here simply that several experiments have been addressed to the question of how effectively decision makers use knowledge of data quality. In most such studies the performance of subjects has been compared with that of some ideal (usually Bayesian) model (see, for examples, Funaro, 1974: Johnson, 1974; Schum, DuCharme, & DePitts, 1973; Snapper & Fryback, 1971; Steiger & Gettys, 1972). What is most germane to the topic of this section is the fact that the models that are used to represent optimal behavior typically distinguish

two separate steps. The first step entails an adjustment of the nominal diagnostic value of a datum, the value that the datum would have if it were known to have been reliably observed or reported. The second step involves the application of the modified datum to the hypotheses of interest. The first step is what we are calling data evaluation, and it is important to note that the failure of subjects to perform this step properly appears to be one of the reasons why they typically acquire less information from data that are not perfectly reliable than is there to be acquired.

5.2 Studies of Data Evaluation

Data evaluation has been recognized by the U.S. Army as being of sufficient importance to warrant the development of a rating procedure for use by tactical intelligence personnel to evaluate all incoming "spot reports" (Combat Intelligence Field Manual, FM30-5). The procedure, which has been standardized for use by NATO army forces, requires that a sender of a report explicitly rate the report both with respect to the reliability of its source and the accuracy of its contents. The letters A through F are used to designate estimates of reliability, and the numbers 1 through 6 to represent judged accuracy. The first five ratings represent a scale going from "completely reliable" (A) to "unreliable" (E) in one case, and from "confirmed by other sources" (1) to "improbable" (5) in the other. The lowest rating in each case is used to indicate that a judgment cannot be made: "reliability cannot be judged" (F), "truth cannot be judged" (6).

Obviously, the purpose of using such a rating procedure is to provide the receiver of a report with some indication of how much confidence he should have in its contents. How effective the procedure has been, however, is open to question. Data collected during field exercises have indicated that ratings often are omitted from spot reports, and that the ratings that are used are too consistently high (Baker, McKendry, & Mace, 1968). The same study also revealed that the reliability and the accuracy ratings tend to be highly correlated. One possible explanation of this correlation is that reliable sources tend to produce accurate reports. This is an intuitively plausible explanation, and it raises the question of the need for two ratings. The other possible explanation for the correlation is that the rater finds it difficult to treat reliability and accuracy as independent dimensions. The results of a subsequent laboratory study of rating behavior were interpreted as supporting the latter possibility (Samet, 1975a). On the basis of his results, Samet proposed that an attempt be made to design and validate an improved procedure for evaluating intelligence data. Specifically, he suggested the possibility of assigning to a report a single number that would represent the evaluator's estimate of the likelihood of the report being true, based on all the information available to him that was relevant to that judgment.

5.3 The Use of Nonquantitative Qualifiers

Probably most people who evaluate data or data sources do not do so according to a formal procedure or in quantitative terms. More typically, they use such qualifiers as "usually reliable," "not very dependable," "prone to exaggerations," "very precise," "a bit careless," "very likely," "a rough estimate," and so forth. Such phrases are certainly meaningful and undoubtedly can convey important qualifying information. The problem is that not all people mean the same thing when they use one of these phrases, and what complicates matters is the fact that even a given individual may use the same term to mean somewhat different things at different times.

A number of efforts have been made to measure the extent of agreement between individuals in their use of such qualifying terms. A common experimental paradigm is that of providing subjects with lists of terms or phrases and requiring them to translate the degree of certainty or uncertainty denoted into a numeric (typically probabilistic) estimate. The variance observed among and within subjects in the translation then provides a measurement of agreement. Results of these studies (see, for example, Lichtenstein & Newman, 1967; Johnson, 1973; Samet, 1975a, 1975b) typically show very low levels of agreement among subjects, and the potential for considerable misunderstanding when large vocabularies of qualifiers are used.

What factors influence the translation of a qualifier into a numeric estimate? There seem to be no clear answers to this question. Cohen, Dearnley, and Hansel (1958) suggested that context in which a word is used might play a role, but a recent study by Johnson (1973) in which the encoding of 15 different probability words (or phrases) contained in each of three different sentence contexts was explored failed to uncover any significant context effect. On the other hand, a study by Rigby and Swain (1971) in which magnitude-denoting terms such as "couple," "lots," and "bunch" were used did suggest such an effect. For example, a "bunch of missiles" had an average assignment of 7.73, while a "bunch of tents" had an average assignment of 12.32. It seems obvious on the face of it that nonquantitative terms denoting physical magnitudes must be subject to enormous context effects. "Small" distances are measured in angstrom units by nuclear physicists and in light years by astronomers. Indeed, it is difficult to see how, in the absence of context, such terms can be considered meaningful at all. Probability terms are different from magnitude terms in that probabilities are bounded whereas magnitudes are not. Perhaps this helps to account for the former's greater independence of context. It should be noted that neither Johnson nor Rigby and Swain found significant differences in the use of these terms due to group membership (army enlisted men and

graduate students, in the former study; army helicopter, Air Force prop, Air Force jet, and Navy attack bomber pilots in the latter).

5.4 Data Evaluation and Training

It seems clear that full exploitation of computer-based tactical data-analysis systems will ultimately require the use of numeric values in place of qualitative estimates, if the reliability of data is to be taken into account when they are used. How best to arrive at these values is, at this point, a matter of conjecture. One could attempt to establish a formal vocabulary of qualitative terms and phrases, associate with each term or phrase a specific numerical value (or range of values), and train personnel to use the resultant isomorphisms in encoding and decoding communications. This is the essence of a proposal made some years ago by Kent (see Platt, 1957). Considering, however, that formal training would be a requirement in any case, a preferred alternative to this approach is to instruct decision makers in the use of probability (and magnitude) scales and require estimates to be communicated in explicitly quantitative terms (Johnson, 1973; Samet, 1975). The obvious problem for training research is that of developing effective procedures for training people to evaluate data quantitatively and for increasing the intra- and inter-person consistency with which quantitative assessments are made.

SECTION VI

PROBLEM STRUCTURING

An exceedingly important step in solving any problem is to be quite explicit about what the problem is that one is to solve. And one way to be explicit is to attempt to represent the problem in terms of a formal structure. While the need to be explicit may appear to be too obvious to deserve comment, it is also apparent that satisfying that need is not always an easy thing to do. Attempts to apply computers to problem-solving tasks have highlighted both the need for explicitness and the difficulty of obtaining it. Armer (1964) has commented on the frustration that is sometimes entailed when one tries to formulate a problem in such a way that a computer can help solve it. He illustrates his point with reference to a bank official who stated, after having his banking procedures mechanized "that 65 percent of the data-processing group's effort went to deciding in detail what problem they were solving" (p. 250). Presumably, the investment was worth it; without it, they could not have recognized a solution had they found one.

The act of trying to make the structure of a problem explicit can be an instructive experience for a problem solver, inasmuch as it forces on him the realization of what he does and does not know about the problem on which he is working--or thinks he is. Essentially, this observation is made by Cloot (1968) vis-a-vis the application of computers to the decision problems of management. He takes the position that one of the major benefits that is to be derived from an attempt to implement a computer-based management information system is not the help that one would get from a functioning system, but what one can learn about the practice of management from the implementation effort. "It can even be argued that the successful use of a computer-based MIS should be measured by the extent to which managers learn to improve their performance so that they can discard it again... There is no doubt that the changes that do come about will be due more to managers having a better understanding of their decision processes than to the technical facilities of the computer" (p. 280).

A major contribution of theoretical treatments of decision making is the provision of formal models in terms of which a decision maker can attempt to structure his own decision problems. Invariably, such models are simplified abstractions, and consequently may not do justice to the full details of any given situation. Nevertheless, they do provide one with structured ways of viewing things, which may make the problems easier to think about, and as a consequence--hopefully--easier to solve. It has been suggested that this is the way in which quantitative models will have their primary effect: "I believe that the greatest impact of the quantitative approach will not be in the area of problem solving, although it will have growing usefulness there. Its greatest

impact will be on problem formulation: the way managers think about their problems--how they size them up, bring new insights to bear on them, and gather information for analyzing them. In this sense, the results that 'quantitative people' have produced are beginning to contribute in a really significant way to the art of management" (Hayes, 1969, p. 108).

6.1 State-Action Matrices

Probably the most well-known way of representing decision situations is in terms of state-action matrices. Such matrices make three aspects of decision situations explicit: the hypothesized possible "states of the world," the action alternatives that are open to the decision maker, and the decision maker's preferences with respect to the various possible state-action combinations. Sometimes such matrices are referred to as payoff matrices inasmuch as each cell of the matrix represents the cost or value--or utility --to the decision maker of the outcome of a particular action selection, given that the associated state hypothesis is true. A decision problem may be represented in this way as follows:

		Action Alternatives					
		A_1	A_2	...	A_j	...	A_n
Hypothesized	H_1	U_{11}	U_{12}				
States							
of the	H_2	U_{21}	U_{22}				
World							
	⋮						
	H_i				U_{ij}		
	⋮						
	H_m						U_{mn}

Much of the theoretical-analytical work on decision making has been concerned with optimal strategies for selecting action alternatives once the situation has been formally structured. Given an explicit decision goal (e.g., minimization of risk, maximization of expected gain, and a formal representation of the situation, prescriptive models can provide useful guidance for action selection. The process of representing real-life decision situations formally, however, is at the present time more of an art than a science. Examples of decision situations that are easily structured can always be found; however, not all decision problems can readily be forced to fit the same mold.

Even given that the structure shown above is an appropriate one for a particular problem, it is clear that in order to use it one must be able to specify, as a minimum, what the hypothesized states of the world are, what one's action options are, and how the various possible decision outcomes (state-action pairs) relate to the value system that will determine the desirability of the actual outcome. One may find it necessary to engage in a considerable amount of information seeking in order to fill out such a structure. Moreover, how one does fill out the structure is determined in part by exogenous variables over which one has no control, and in part by self-imposed constraints. The state of the world tends to be beyond one's control; all one can do is attempt to determine what it is likely to be. One's action alternatives, however, may be constrained in part by limits that are self-imposed. What are viewed as viable strategic military options, for example, may depend on the particular military doctrine in vogue at the time. Benington (1964) points out that the basic concept behind the development of such automated, or semiautomated, systems as the SAGE system in the 1950's was the concept of "set-piece warfare." "Set-piece warfare is characterized by warning of threat, total and preplanned goals, speed of response, and detailed and precise management of the campaign" (p. 9). Emphasis is on massive retaliation totally preplanned, or "spasm" response. During the early 1960's, the set-piece warfare idea lost favor. President Kennedy and Secretary of Defense McNamara began to emphasize the importance of flexibility and adaptability, the ability to make selected and controlled responses, directed toward military (non-civilian) targets and appropriate to the (not always foreseeable) contingencies that elicit them. Clearly, the set of action alternatives that the strategist will consider under one of these retaliation doctrines is quite different from that that he will consider under the other.

6.2 Alternative Structurings of a Given Situation

It is apparent that to think in terms of the structure of a decision space is to oversimplify matters greatly. Usually any given situation can be structured in a variety of ways. Moreover, how one chooses to represent a particular situation may not be incidental. It seems to be true of problem solving in general that how one represents a problem can be an important factor in determining how easily one can then solve it. This point has often been made by individuals engaged in efforts to program computers to perform intellectually demanding tasks (see, for example, Nilsson, 1971). The same problem may yield to attempts to solve it when represented in one way while resisting such attempts when represented in another.

An important aspect of developing a useful structure is that of conceptualizing a situation at an appropriate level of detail.

Too simple a structure may violate the complexity of an actual situation. On the other hand, Charles Pierce's maxim that "a few clear ideas are worth more than many confused ones" seems particularly apropos here. We state as a conjecture that a necessary requisite for effective decision making is the ability to get quickly to the heart of a problem, to concentrate on essentials, and to ignore irrelevancies. What this often means in practice is being able to see through superficialities that frequently obscure underlying issues. Moreover, even when the situation, stripped of incidentals, is inherently complex, there may be some merit in a simplified conceptualization of it, provided that the fact that the conceptualization is a simplification is not then promptly forgotten. There is little to be gained by representing a situation in such a complex way that the decision maker cannot grasp the representation intellectually. What constitutes an optimal level of detail may vary from situation to situation and from individual to individual, but variability in this regard may not be very great. We suspect that for the vast majority of situations and decision makers a representation that involves more than eight or ten hypothesized states of the world and as many action alternatives, at any given level of description, will prove to be an unwieldy one.

6.3 Structuring as an Iterative Process

On the basis of an anlaysis of protocols obtained in his classical study of problem solving, Duncker (1945) reached a conclusion that is germane to the issue of problem structuring. The problem that he used most frequently in his studies was the now well-known radiation problem: "given a human being with an inoperable stomach tumor, and rays which destroy organic tissue at sufficient intensity, by what procedure can one free him of the tumor by these rays and at the same time avoid destroying the healthy tissue which surrounds it?" (p. 28). The conclusion that Duncker came to after observing the efforts of many people to solve such problems was that the development of a solution typically proceeds from the more general to the more specific. (On this point, see also Hogarth, 1974, and Kleinmutz, 1968.) The principle by which the problem is, hopefully, to be solved emerges first, and the details of the solution come later. It often happens that a principle may be valid, but there turns out to be no feasible way to implement it. A principle that was frequently identified in the case of the radiation problem, for example, was "avoid contact between rays and healthy tissue." When the problem solver could think of no way to do this and still get the rays to the tumor, he had to abandon the principle itself--even though it was a sound one--and search for another that was not only sound but practicable.

The finding of a new principle, or a general property of a solution, always involves, Duncker suggests, a reformulation of the original problem. In the case of the example just given, having accepted "avoiding contact" as a valid principle, one has in effect

defined his problem as that of finding a way to do just this. When forced to reject a given principle as impractical, the substitution of another (e.g., "lower the intensity of the rays on their way through healthy tissue") in effect defines another how-to-do-it problem to be solved. "We can accordingly describe a process of solution either as development of the solution or as development of the problem. Every solution-principle found in the process, which is itself not yet ripe for concrete realization... functions from then on as reformulation, as sharpening of the original setting of the problem. *It is therefore meaningful to say that what is really done in any solution of problems consists in formulating the problem more productively.* To sum up: *The final form of a solution is typically attained by way of mediating phases of the process, of which each one, in retrospect, possesses the character of a solution, and, in prospect, that of a problem*" (p. 34, italics his).

It is probably the case that complex decision problems, like other types of complex problems, yield grudgingly to attempts to structure them. Moreover, a decision maker may find it necessary to formulate and reformulate a decision space several times before arriving at a structure that he feels adequately represents the decision problem that he must solve and does so in a way that facilitates arriving at a solution. The willingness to discard a favored conceptual framework when it is seen no longer to fit the facts in hand has been considered by some to be one of the defining characteristics of original thinking (Mackworth, 1965; Polyani, 1963).

6.4 Problem Structuring and Training

The question of how to train decision makers to structure decision problems effectively has received very little attention. Moreover, if it is true, as Edwards (1973) has suggested, that of the several aspects of decision analysis the process of problem structuring is least amenable to formal prescription, exactly what should be taught is not clear.

It seems likely, however, that something is to be gained by familiarizing decision makers with such formal representations--models--of decision situations, as are provided by decision theory and game theory. Such training should be conducted in such a way as not to leave the student with the unrealistic idea that all decision situations are readily represented--without distortion--by the same model.

Practice in representing specific situations in terms of such models, and criteria for judging the relative merits of different models for different problems should probably be part of any training program in decision making. Practice in representing a given decision problem at different levels of detail also would

probably be beneficial. Duncker's work suggests that one approach to problem structuring that might usefully be taught is that of zeroing in on an appropriate formulation by a series of approximations, proceeding from the more general to the more detailed.

But these are only conjectures. The fact is that little is known about how to train a person to be good at imposing structure on a problem--whether it be a decision problem or a problem of any other kind. Mackworth (1965) has noted that one of the characteristics of creative individuals is an exceptionally strong need to find order where none appears on the surface. If this is so, then one way to train people to be better problem structurers is to train them to be more creative. If only we knew how to do that!

An alternative to training decision makers to formalize their decision problems is to provide them with models that are appropriate to their particular situations, and that can then be used as decision aids. Gorry (1970) has suggested this possibility. A model that is to be used by a decision maker need not be generated by him, but, Gorry points out, it may be derived from his description of the situation, and it must be thoroughly understandable by him. In this case the training task becomes that of teaching an individual to make effective use of the structure that someone else has imposed upon his problem.

At least one study has been addressed to the question of the subtasks in terms of which one class of decision makers sees decision making and how this view would change as a result of training. Hill and Martin (1971) gave secondary-school teachers problem-solving exercises designed to train them with respect to some of nineteen specific skills that they associated with decision making and to acquaint them with a particular model of the decision-making process (see Section III). Both before and after training, the subjects were asked to list the specific steps that they would take in an effort to solve a hypothetical problem involving an interperson conflict. Perhaps the most striking aspect of the results was how large a proportion of the steps that subjects listed tell in the "formulating-action-alternatives" category. Before training, more of the listed steps fell in this category than in the other five combined. The main effect of training was to reduce the number of steps in this category by about two-thirds and to increase the usage of some of the other categories slightly; but formulating alternatives still remained the largest category. The investigators concluded that training had made the participants more aware of the several activities involved in decision making, but pointed out that their results shed no light on the question of whether much as increased awareness would produce better decision making.

SECTION VII

HYPOTHESIS GENERATION

Hypothesis generation is closely associated with problem structuring. We find it convenient to consider it separately, however, because it is a more narrowly focused type of activity. Problem structuring is always important. Even when complete information is available concerning the state of the world, the action alternatives and all the possible decision outcomes, it is still necessary to cast the problem into some mold, and the mold that is chosen may have much to do with the decision that is made. Hypothesis generation, on the other hand, is a necessary activity in those decision situations characterized by uncertainty about such things as the state of the world and the implications of selecting specific decision alternatives. Often, in spite of one's best efforts to gather information, it is not possible to eliminate uncertainty about these things completely. In such cases, it is convenient to conceptualize the decision maker's view of the situation as a set of conjectures, or hypotheses.

7.1 Hypothesis Generation versus Hypothesis Testing

Investigators of cognitive processes have long recognized two rather different types of thinking. Bartlett (1958) speaks of closed versus adventurous thinking, Guilford (1963) of convergent versus divergent thought. Mackworth (1965) distinguishes problem solvers and problem finders. The one kind of thinking tends to be deductive and analytical; the other inductive and analogical. The first has to do with evaluating hypotheses, the second with generating them. The history of science attests to the fact that the ability to evaluate hypotheses, to deduce the implications of theories and put them to empirical test, is a far more common quality among men than is the ability to generate hypotheses, to construct theories that organize and structure facts that were not perceived as related before.

Some formal treatments of decision making require that the situation, as viewed by the decision maker, be conceptualized as a set of mutually exclusive and exhaustive hypotheses, each of which represents one of the possible states of the world. As data are gathered, they are used to modify a set of probabilities, each of which represents the decision maker's estimate of the likelihood that a given hypothesis is true. Much laboratory experimentation has been devoted to the question of how effectively man can assimilate data and use it to modify his view of the world as implied by the probabilities that he associates with the hypotheses that he is entertaining. (We will consider that problem in the following section.) However, very little attention has been given to the question of how capable people are of generating a reasonable set of hypotheses to begin with, or of modifying the set when the need to do so arises.

Typically, all of the hypotheses that are to be considered are provided for the decision maker in advance, so the process of hypothesis generation is not studied. Moreover, formal decision procedures usually permit the decision maker <u>only</u> to update the probabilities that have been assigned to the previously established set of hypotheses. They fail to recognize the fact that it may be the case in real-life situations that a set of hypotheses that is originally developed may not contain the hypothesis that will eventually prove to be the true one. It often occurs in real-life situations that incoming data suggest to the decision maker new hypotheses that have not yet been considered. Any decision-making procedure that purports to be generally valid must provide for establishment of new hypotheses whenever the information in hand indicates the need for them.

7.2 Importance of Hypothesis Generation

The importance of the function of hypothesis generation can hardly be overemphasized. To be sure, one may think of some decision contexts for which all the potentially interesting hypotheses can be specified in advance. For example, it may be the case for some straightforward troubleshooting situations that an exhaustive set of the hypotheses of interest can be listed prior to the performance of any tests. More typical of complex decision problems, however, is the case in which the set of possibilities is either not fully known, or too large to be listed exhaustively. The problem of the physician who is attempting to diagnose an illness with a set of symptoms that does not fit a common pattern, or the investor who is trying to gauge the risks and potential gains in a speculative financial venture, or the computer programmer who is tracking down an elusive bug, or the tactician who is trying to assess the significance of some unorthodox behavior on the part of a wily opponent is less that of testing prespecified hypotheses than that of defining hypotheses that it would make sense to consider.

The difficulty is not so much that of representing a decision situation in terms of a set of possible states of the world that is exhaustive and mutually exclusive. The problem is that of coming up with a set of possibilities that is <u>useful</u> from the decision maker's point of view. A military commander can always represent the alternatives that are open to an adversary in terms of such gross action categories as attack, defend, and withdraw, and the ability to distinguish among these possibilities would undoubtedly be of interest. However, a commander's decision-making responsibilities typically require much more precise information than would be provided by the resolution of the uncertainty implicit in these three possibilities. That is to say, he wants to know not only whether enemy forces plan to attack, but at what time, in what strength, at what locations, and so forth. It is at this level of representation that the commander's (or perhaps his intelligence officer's) hypothesis-generation capabilities are put to the test.

7.3 Experiments on Hypothesis Generation

The study of hypothesis generation in the laboratory has often involved "concept attainment" or "discover the rule" type tasks. The work of Bruner, Goodnow, and Austin (1956) illustrates the use of concept attainment tasks to study this aspect of thinking. In a typical experiment, a subject attempts to identify a concept that an experimenter has in mind. The concept usually is defined in terms of conjunctions or disjunctions of specific stimulus attributes (e.g., "red and square"; "blue or yellow, and not circular"). In some situations the subject is shown stimuli, some of which belong to the conceptual category that he is attempting to identify and some of which do not. He is told which stimuli are which and from this "exemplar" information he is to attempt to identify the concept. Sometimes the subject chooses the stimuli that he sees, in which case the task can also be used to study a form of information-gathering behavior.

Obviously, the performance of this task involves hypothesis testing (a topic to which we will turn in the following section), but the key problem is that of hypothesis generation. Unless one comes up with the right hypothesis to test, the testing that he does will only eliminate some of the untenable possibilities, of which there may be many.

A basic conclusion that Bruner et al. draw from their experimental results is that the strategies that subjects employ in these sorts of tasks can be isolated and described. They identify four such strategies, for example, that subjects use when they have the job of discovering a conjunctive concept by selecting stimuli and being told, concerning each stimulus selected, whether or not it is an exemplar of the concept that they are attempting to identify. These strategies differ in terms of the balance they strike among three parameters: the amount of information obtained from an observation, the cognitive strain imposed on the subject (amount of information that must be carried in memory, extent to which involved inferences must be made), and the risk that the strategy will fail. The strategies are defined in terms of the nature of the hypotheses that are generated and put to the test. In one case, for example ("successive scanning"), one specific concept is hypothesized at a time, and stimuli are chosen in such a way as to test that hypothesis directly. In another case ("conservative focusing"), the initial hypothesis, in effect, includes several possible concepts and an attempt is made to discover the defining attributes systematically one at a time. Which of the several strategies is most appropriate depends on the details of the experimental situation.

Bruner et al. found that the strategies that subjects use tend to change appropriately in response to changes in the

experimental situation; and, on balance, these investigators considered the performance of their subjects to be quite good. In their words: "In general, we are struck by the notable flexibility and intelligence of our subjects in adapting their strategies to the information, capacity, and risk requirements we have imposed on them. They have altered their strategies to take into account the increased difficulty of the problems being tackled, choosing methods of information gathering that were abstractly less than ideal but that lightened pressures imposed on them by the tasks set them. They have changed from safe-but-slow to risky-but-fast strategies in the light of the number of moves allowed them. They have shown themselves able to adapt to cues that were less than perfect in validity and have shown good judgment in dealing with various kinds of payoff matrices. They have shown an ability to combine partially valid cues and to resolve conflicting cues" (p. 238).

Performance was not ideal, however. Among the limitations that were noted were a tendency to persist in focusing on cues that had proved to be useful in the past even if they were not useful in the present, and an inability to make as effective use of information gained from noninstances of a category as of that gained from category exemplars.

Bruner et al. also found that concepts defined in terms of disjunctions of stimulus attributes were more difficult to discover than those that were conjunctively defined. This finding has been corroborated by Neisser and Weene (1962) who used a large variety of attribute-combination rules. Not surprisingly, concepts defined in terms of the presence or absence of a single attribute are easier to attain than are those defined in terms of conjunctions or disjunctions of two or more attributes, which in turn are easier than those defined in terms of more complex rules involving combinations of conjunctions and/or disjunctions (Haygood & Bourne, 1965; Neisser & Weene, 1962).

Another experimental task that has been used to study hypothesis generation is that of discovering the rule by which a specific sequence of numbers or letters was generated. Typically, the subject is shown one or more sequences (or segments of sequences) that satisfy the rule. He then can propose other sequences, or continuations of the segment, in order to test the validity of tentative hypotheses that he may wish to consider. Each time he proposes a possibility he is told whether it satisfies the rule; and when he feels he has obtained enough information to justify doing so, he is to state the rule.

Again, performance of this task obviously involves information gathering and hypothesis testing as well as hypothesis generation, but hypothesis generation is in some sense central. What information

is sought is likely to depend strongly on what rule is being considered. Moreover, unless the correct rule is hypothesized at some point, it cannot be tested and validated.

The results of experiments along these lines have revealed some interesting deficiencies in hypothesis-generation behavior which appear to stem from a lack of understanding of some basic rules of logic. Wason (1974) has described some results that suggest that people may have particular difficulty in discovering rules that are sufficiently general that they subsume many rules that are more specific. For example, the rule "any three numbers in increasing order of magnitude" proved to be particularly difficult for his subjects to discover. If, as examples of triads that conform to this rule, a subject were given (8 10 12), (14 16 18) and (20 22 24), he might quickly generate the hypothesis "successive even numbers," test it with other sequences that satisfy it, and then announce this rule with confidence. What is disappointing about this behavior is the failure to hypothesize alternative rules to which the given sequences also conform, and then to consider sequences that would discriminate between the alternatives hypothesized. More disturbing, however, is the finding that even when told of the incorrectness of a hypothesis, and presented with conclusive infirm evidence, subjects sometimes insisted that their hypothesized rule was validated by the fact that all the test sequences that they generated conformed to it.

Two other results noted by Wason are relevant to the problem of hypothesis generation, because they also demonstrate how the process can get bogged down. First is the possibility of perseveration with an invalidated hypothesis without recognizing that one is perseverating. He notes, in this regard, that what subjects often do when informed that a hypothesized rule is not the correct one is to generate additional triads that are consistent with that rule and then announce the same rule expressed in different terms. Second is a tendency, when hypothesized rules are invalidated, to generate more and more complex rules rather than simpler ones. The following example is given of a third generation rule produced by one subject: "The rule is that the second number is random, and either the first number equals the second minus two, and the third is random but greater than the second; or the third number equals the second plus two, and the first is random but less than the second" (p. 382). Recall that the correct rule was "any three numbers in increasing order of magnitude." One conclusion that may be drawn from this type of experimental finding is that the discovery of a general rule, even though conceptually simple, may be impeded by the discovery of more specific rules whose exemplars are also exemplars of the more general rule.

7.4 Hypothesis Generation and Training

Hypothesis generation represents the same sort of challenge to training and training research as does problem structuring. The basic need in both cases is for a greater understanding of how to promote creative thinking.

A specific problem that deserves attention from training specialists is that of perseveration. Results such as those obtained by Bruner, Goodnow, and Austin (1956) and by Wason (1974) indicate the need for training procedures designed to improve the ability, or increase the willingness, of decision makers to generate alternatives to the hypothesis, or hypotheses, under consideration. They demonstrate the importance of sensitizing decision makers to the danger of accepting a hypothesis on the basis of insufficient evidence, and to the fact that the best way to avoid this mistake is to attempt to generate plausible alternatives and to seek the kind of data that will be most likely to discriminate among them.

SECTION VIII

HYPOTHESIS EVALUATION

Narrowly defined, hypothesis evaluation refers to the process of applying data to the assessment of the likelihoods of one's hypotheses concerning the unknowns of the situation. More generaly, the term might be used to connote the process of extracting information from data, of attempting to reduce one's degree of uncertainty about the parameters of the decision space. In some formally structured approaches to decision making, hypothesis evaluation may involve the revisions of numerical probability estimates or other quantitative indicants of relative likelihoods. In other cases the process may be less explicit, but it is not for that reason less important. We assume that even in situations that have been given little formal structure, the decision maker attempts to make use of at least some of the data that are available to him, in order to clarify his view, or perhaps to confirm his assessment, of the situation.

The following discussion takes a rather broad view of hypothesis evaluation. It touches on a number of topics that relate to man's abilities, limitations, biases and predilections as a processor of information or a user of evidence. In some cases it may appear to range beyond the specific subject of hypothesis evaluation, and deal with "thinking" more generally. Our reason for including this material is that it seems to us relevant to the problem of decision making, and it appears to fit more readily here than elsewhere within our conceptual framework. In Section 8.6, the discussion becomes narrowly focused on the problem of revising probabilities in situations that have been formalized to the extent that a Bayesian data-aggregation algorithm might be applied.

8.1 Serial versus Parallel Processing

One question of interest concerning the way people evaluate hypotheses is whether they consider them one, or several, at a time. Empirical data are lacking on the question of which of these alternatives best characterizes man's approach to hypothesis evaluation. It is our impression that the prevailing consensus is that the assumption of seriality is the more plausible of the two, insofar as the conscious consideration of hypotheses is concerned.

If the serial model is the more nearly correct, this must represent a basic limitation of man. It is difficult to think of a convincing reason why one should evaluate the hypotheses serially if he is able to treat them in parallel.

But even if we assume that one cannot test several hypotheses at once, there is still a question about the order in which testing is done. One might apply an incoming datum to each of the

hypotheses in turn. Alternatively, one might focus exclusively on one hypothesis until one had enough confirming data to accept it, or until the evidence against it was sufficient to warrant its rejection, in which case attention would be shifted to another possibility. Note that in this latter case a datum cannot be discarded after being applied to the evaluation of one hypothesis because it may be germane to the evaluation of others later.

One putative advantage of the Bayesian approach (see Section 8.6) is that it forces the decision maker to apply an incoming datum to each of the candidate hypotheses in turn. One of the implications of this fact is that it minimizes the need for the decision maker or system to retain data. Assuming that the set of hypotheses with which the decision maker is working is complete, and will not be extended, a datum can be discarded once it has been assimilated and the probabilities associated with all the hypotheses revised.

8.2 Subconscious Processes

What is happening at a subsconscious level is, of course, even less well-understood. The belief has been expressed that the brain carries on problem-solving activity even when one is not consciously thinking about a problem. Wallas (1926) elaborated and popularized the notion, which he credits to Helmholtz, that creative thinking often involves a period of "incubation," which follows a period of "preparation," and precedes a period of "illumination." During the preparation period, according to this view, the problem solver consciously labors on the problem,; during the illumination period the problem solver becomes aware of the solution for which he was seeking. No conscious attention is given to the problem during the incubation period, but, Wallace suggests, much subsconscious exploration of the problem takes place.

While the idea has primarily anecdotal support, the testimony of creative thinkers about the way they have arrived at solutions to difficult problems is fairly compelling evidence that something of this sort does occur. We mention it in this context to make the point that the fact (if it is a fact) that decision makers tend to apply newly acquired data to the evaluation of only one hypothesis at a time, should probably not be taken as conclusive evidence that the credibility of a hypothesis not under consideration has not been affected by those data. Moreover, it is at least a plausible conjecture that the likelihood that any given hypothesis will"suggest itself" for explicit consideration may depend to some degree on such subconscious activity (Maier, 1931).

Dreyfus (1961) has argued that such subconscious, or marginally conscious, activity is a general and difficult-to-simulate characteristic of man as a problem solver. It is this ability that makes it possible for him to consider consciously only the "interesting" moves in a game of chess without explicitly considering all possible moves and rejecting those that are not worth pursuing. But subconscious processes are beyond the scope of this report, so we will not pursue the topic further.

8.3 Man As An Intuitive Logician

Technically, logic is the discipline which deals with the rules of valid inference. The term is used colloquially, however, as a synonym for reasoning. It is of some relevance to the general problem of decision making, and in particular to the problem of training decision makers, to consider whether reasoning as it is practiced by people is logical in the technical sense; and, to the extent that it is illogical, whether it is illogical in consistent ways. A further question of interest is whether training in formal logic can reasonably be expected to improve decision-making performance.

Philosophers have not been in agreement on the first question. Henle (1962) points out that some of the 19th century writers (e.g., Boole, 1854; Kant, 1885; Mill, 1874) viewed logic as the science of the laws of thought. Some more recent writers (e.g., Cohen, 1944; Russell, 1904; Schiller, 1930) have treated logic as something quite independent of thought processes and to reject the notion that thinking necessarily conforms to logical principles.* A middle-of-the-road view is that thinking sometimes conforms to logical principles--especially when one's explicit purpose is to reason carefully and deductively--and sometimes does not.

---

*A cynic might assert that few arguments are won or lost on logical grounds. Certainly, the alogical strategems that can be applied to arguments are numerous, and perhaps are better learned in the course of normal development than are the rules of inference. The disputatious reader who feels his arsenal of such strategems is deficient is referred to Schopenhauer (no date) who provides a veritable cornicopia of them.

Whether or not thinking is logical may be difficult to determine empirically in any particular case, because the steps by which one arrives at a conclusion usually are not available for observation. As Mill (1874) points out, since "the premises are seldom formally set out,... it is almost always to a certain degree optional in what manner the suppressed link shall be filled up... [A person] has it almost always in his power to make his syllogism good by introducing a false premise; and hence it is scarcely ever possible decidely to affirm that any argument involves a bad syllogism" (p. 560; from Henle, 1962).

Individuals undoubtedly differ greatly in their ability to think logically, and any characterization of human strengths and weaknesses in this regard is bound to be only partially correct. There are many ways in which reasoning can be illogical, however, and it is not unreasonable to ask whether some of the many possible evidences of fallibility are appreciably more common than others. Several ways in which human reasoning does seem to depart from the ranks of logic have been discussed by Henle (1962). These include: failure to distinguish between the factual truth of a conclusion and the logical validity of the argument on which it is based; restatement of a premise or a conclusion, which may have the effect of preserving a logically valid form, while changing the substance of the argument; the omission of premises from an argument, or the addition of spurious premises. The fallacy of the "undistributed middle" is one that has long been recognized as particularly bothersome, and involves the assignment of different meanings to the same term when it appears in different premises.

Another type of logical error that seems to be commonly made involves a misunderstanding of the syllogistic form: "If A then B; A; therefore B," or "If A then B; not B; therefore not A." These forms may be perverted either as "If A then B; not A; therefore not B," or "If A then B; B; therefore A." Both of these forms are invalid; nevertheless most readers will probably recognize them as forms that they have encountered, and perhaps used, in arguments.

Wason (1974) describes a failure in reasoning that he has observed that seems to be related to this type of misunderstanding. Four cards are placed on a table so the subject can see only one side of each of them. The cards contain respectively a vowel, a consonant, an even number and an odd number. The subject is told that each card has a letter on one side and a number on the other, and is asked which cards would have to be turned over to determine the truth or falsity of the statement: "If a card has a vowel on one side, then it has an even number on the other." The majority of Wason's subjects chose either the card showing the vowel and the one showing the even number, or just the card showing the vowel. The correct answer is: the card that shows the vowel and the one that shows the odd number. Only by finding an odd number behind

the vowel or a vowel behind the odd number would the statement be falsified. The students' choice of the card with the even number is a form of the fallacy known as asserting the consequent: "If A then B; B; therefore A."

This type of reasoning error occurs with sufficient consistency (at least among college students) to have prompted investigation by several researchers. A completely satisfactory explanation has not yet been forthcoming. Wason seems to favor the view that the choice of cards is made on an intuitive basis and that the "reasons" for the choice - which subjects give in response to the experimenter's inquiries - are really rationalizations. "This hypothesis is consistent with our crude knowledge about intuition. A verdict may occur to a judge before the grounds which support it have been spelled out; a chess player may "see" a good move, and then analyze the continuations which validate it. Such thought suggests a processing mechanism which operates at different levels" (p.385).

The last chapter on the topic of the relationship between logic and thought has not been written. And it cannot be until much more is known about the workings of the human mind. The immediate challenge for training research is to identify ways to improve the capability of individuals to reason logically, or at least to recognize and be able to avoid the more common illogical pitfalls.

8.4 Man as an Intuitive Statistician

It is quite clear that most individuals could manage to get through life without ever explicitly assigning a numerical probability to an event. Undoubtedly, the vast majority of people do so. It seems safe to assume, however, that people do make judgments of likelihoods, and that these judgments--even though nonnumeric, and often implicit--condition their behavior. An individual carries an umbrella because he thinks there is a good chance of rain, or buys stock that he expects to appreciate. One purchases life insurance before boarding an airplane because one, in effect, has considered the likelihood that the plane will go down during that flight to be nonnegligible; the fact that he boards the plane at all is probably evidence that he also considers that likelihood to be something less than certainty. One chooses one among three job opportunities, because the chances of success and advancement are perceived as greater in the case of the selected job than in that of the others. In short, although most of us do not attempt to assign numeric probabilities to possible situations or events, we behave as though our choices had been dictated by reasoning of the sort: this event is more likely than that, or the likelihood of this situation is great enough so that I had better do thus and so in order to be prepared if it should occur.

A question of some practical interest, therefore, is that of how effectively such judgments are made. For many situations, there is no way to answer this question objectively. The individual who selects one job from among three possibilities because he considers the likelihood of success to be highest for that case will never know for certain whether his judgment was correct. There are also, however, many situations for which the "objective" probabilities of events are known or can be determined, and we can at least ask how well people do when asked to estimate probabilities explicitly in these cases. The literature that is relevant to this question falls fairly naturally into three categories. First are the studies that deal with people's ability to estimate the statistical properties of samples that they are permitted to observe. Such studies concern relative frequencies rather than probabilities, but to the degree that our ideas about probabilities are based on, or influenced by, perceived frequencies they are germane. Second are some studies that have to do with the extent to which people's intuitive notions about the probabilities of events correspond to, or conflict with, the implications of the theory of probability as represented in the probability calculus. Third are numerous recent experiments that consider the specific question of how effectively people function as Bayesian data aggregators. In this section we will consider briefly the first of these three categories of studies; in Sections 8.4 and 8.5 we will consider the last two.

People appear to be reasonably good at perceiving proportions, or the relative frequencies of occurrence, of both sequential and simultaneous events (Attneave, 1953; Peterson & Beach, 1967; Schrenk & Kanarick, 1967; Erlich, 1964; Vlek, 1970) and at estimating the means of number sequences (Beach & Swensson, 1966; Edwards, 1967). Inferences concerning the median or mode of a skewed distribution (assuming the subject knows the definitions of these terms) are fairly accurate, and the estimated mean of such distributions tends to be biased in the direction of the median (Peterson & Beach, 1967). One's confidence is one's estimate of the mean or the variance of a population appears to increase as the sample size increases (Peterson & Beach, 1967; but see also Pitz, (1967).

Estimates of the variability of a set of data often tend to decrease as the mean increases (Hofstatter, 1939; Lathrop, 1967; Peterson & Beach, 1967). Peterson and Beach (1967) point out that while the notion that variability is necessarily inversely related to the mean is erroneous, it is intuitively compelling. "Think of the top of a forest. The tree tops seem to form a fairly smooth surface, considering that the tree may be 60 or 70 feet tall. Now, look at your desk top. In all probability it is littered with many objects and if a cloth were thrown over it the surface would seem very bumpy and variable. The forest top is far more variable than the surface of your desk, but not relative to the sizes of the objects being considered" (p. 31). One is led to wonder whether

the finding that estimated variability tends to decrease with increasing mean might be due in part to failure by the subject to understand that it is an estimate of absolute variability that he is to produce. Relative variability probably often does decrease as the mean increases (to cite Peterson and Beach's tree top example), and without explicit instructions to the contrary it would not be unreasonable for a subject to suit the terms to the context, as one does when one speaks both of a small skyscraper and a large dog.

8.5 Intuitive Probability Theory

How closely do man's intuitions about probabilities correspond to the implications of probability theory? The question cannot be answered decisively, but a number of pertinent observations can be made. For example, people often seem to find it difficult to believe that the outcome of an event can be independent of what has preceded it. This difficulty is sometimes manifested in the "gambler's fallacy" (a fallacy that competent gamblers probably would not make), one form of which holds that a run of successes increases the likelihood of a failure, or vice versa (Cohen & Hansel, 1956). Another example of assumed dependence among successive events has been noted by Jarvik (1951), who found that when given a two-alternative prediction task, subjects often tended to predict the more frequent event after one occurrence of the less frequent event and to predict the less frequent after two consecutive occurrences of the more frequent event.

Several experimenters have found that man does not estimate the probability of compound events very accurately. In particular, when assessing the likelihood of the joint occurrence of several independent events, he tends to produce estimates that are too high (Cohen, Chesnick, & Haran, 1972; Fleming, 1970; Slovic, 1969). Conversely, when estimating the probability of disjunctive events--the probability that any one of several specified events will occur--he tends to produce estimates that are too low (Cohen, Chesnick, & Haran, 1972; Tversky & Hahneman, 1974). The overestimation of the probability of conjunctive events is consistent with the observation that people frequently base judgments of the degree of correlation between two events on those cases in which the outcomes of interest do occur together without giving sufficient consideration to those cases in which they do not (Peterson & Beach, 1967).

What is of more interest than the fact that man's intuitions sometimes lead to incorrect judgments about event probabilities is the question of the extent to which the failings of intuition--at least insofar as they are systematic--are explainable in terms of identifiable ways in which such judgments are made. In a recent series of studies. Tversky and Kahneman (1971, 1973, 1974;

Kahneman & Tversky, 1972, 1973) have explored this question. The general approach in these studies and in those of others who have conducted similar investigations (e.g., Alberoni, 1962; Tune, 1964; Wagenaar, 1970) has been to ask people to estimate the probability of the occurrence of a hypothetical event, or, perhaps more commonly, to indicate which of two such events is the more probable. One might be asked, for example, to indicate which of the two following sequences of coin tosses is the more likely, HHHHTTTT or HHTHTTHT; or to indicate which of two hospitals-- which record approximately 15 and 45 births a day, respectively--would have the largest frequency of days on which more than 60% of the babies born are boys.

The results of these studies have revealed a number of ways in which the answers that people give to such questions depart systematically from the objective probabilities of the events as inferred from the application of probability mathematics. Tversky and Kahneman attribute such failures in judment to the heuristic principles that people often use when attempting to estimate probabilities or relative likelihoods.

It will be helpful, before considering some of Tversky and Kahneman's specific results to digress briefly to consider the notion of a heuristic principle or procedure. The term "heuristic," which comes from the Greek heuriskin, meaning "serving to discover," appears sporadically in the literature of philosophy and logic as the name of a branch of study dealing with the methods of inductive reasoning. It was revived by Polya (1957) in his classic treatise on problem solving, and used to connote inductive and analogical reasoning leading to plausible conclusions, as opposed to the deductive developments of rigorous proofs. In recent years, computer scientists, and especially researchers in the area of machine intelligence, have appropriated the term to connote "a rule of thumb, strategy, trick, simplification, or other kind of device which drastically limits search for solutions in large problem spaces" (Feigenbaum & Feldman, 1963, p. 6). In short, a heuristic principle or procedure, usually referred to simply as a heuristic, is a means of making an inherently difficult problem more tractable. The criterion by which a heuristic is measured is its usefulness. It is important to bear in mind, however, that heuristics are not expected to lead invariably to correct solutions. "A 'heuristic program,' to be considered successful, must work well on a variety of problems, and may often be excused if it fails on some" (Minsky, 1963, p. 408).

8.5.1 Representativeness

Tversky and Kahneman describe two heuristic principles--representativeness and availability--which they feel account for many of the systematic judmental biases that they and other investigators have observed. According to the representativeness principle, "the subjective probability of an event, or a sample, is determined by the degree to which it: (i) is similar in essential characteristics to its parent population; and (ii) reflects the salient features of the process by which it is generated" (Kahneman & Tversky, 1972, p. 430). Several examples of the application of this principle are given; two will suffice for our purposes, one illustrating each of the subprinciples.

The importance of the similarity between the judged event and the parent population is illustrated by the following question: "All families of six children in a city were surveyed. In 72 families the exact order of births of boys and girls was GBGBBG. What is your estimate of the number of families surveyed in which the exact order of births was BGBBBB?" (Kahneman & Tversky, 1972, p. 432). If the probabilities of male and female births were exactly equal, the two birth sequences would be equally probable. (Apparently, the frequency of male births is slightly higher than that of female births, so the latter sequence is slightly more probable than the former.) About 80% of the subjects (high-school students) who were asked this question judged the latter sequence to be less likely than the former; the median estimated number of families with this birth order was 30. Kahneman and Tversky attributed this result to the fact that the two birth sequences, while about equally likely, are not equally representative of families in the population. The former sequence is more similar to a larger proportion of the population, both in terms of the relative number of girls and boys, and in terms of the length of runs of births of the same sex.

The second way in which the representativeness heuristic manifests itself--in sensitivity to the degree to which an event reflects the salient features of the process that generated it--is illustrated by the tendency of people to consider regularities in small samples to be inconsistent with the assumption that such samples were generated by a random process. Thus, when people are asked to produce random sequences such as the results of an imagined series of coin tosses, they tend to produce fewer long runs than would a truly random process. Moreover, in judging the randomness of small samples, they are likely to reject as nonrandom many of the samples that a random process does generate. Kahneman and Tversky characterize the intuition that produces such judgmental biases as a belief that a representative sample should represent the essential characteristics of the parent population, not only globally, but locally in each of its parts. In other words the

observed behavior is consistent with the belief that the law of large numbers applies to small numbers as well (Tversky & Kahneman, 1971).

The application of this heuristic could lead one to the sort of fallacious thinking illustrated by the conclusion that the probability of finding more than 600 boys in a random sample of 1000 children is the same as that of finding more than 60 boys in a random sample of 100 children. The probability of the latter event is, of course, much greater than that of the former. Kahneman and Tversky (1972) showed that people (at least high-school students) do virtually ignore the effect of sample size when estimating the probabilities of random events of this sort. In general, the estimates made by Kahneman and Tversky's subjects, when asked to judge the probability of events that have a binomial distribution, were much more appropriate for small samples (e.g., [illegible]) than for large samples (e.g., 100 or 1000). In other words, for large samples, subjects tended to underestimate grossly the probability of high-probability events and overestimate the probability of low-probability events, and the magnitude of the miss increased with the size of the sample.

8.5.2 Availability

The availability principle, according to Tversky and Kahneman (1973) is used whenever one bases estimates of frequency or probability on the ease with which instances or associations are called to mind. For example, when asked to estimate the relative likelihoods of heart attacks for men and women, one might think of male and female victims of heart attack among one's personal acquaintances and take the ratio as an estimate of the relative likelihoods in the population. Or, if asked to judge which of two letters occurs the more frequently as the first letter of English words, one might attempt to think of a few words of each class and make the judgment on the basis of the rapidity with which examples come to mind.

Tversky and Kahneman point out that "availability is an ecologically valid cue for the judgment of frequency because, in general, more frequent events are easier to recall or imagine than infrequent ones. However, availability is also affected by various factors which are unrelated to actual frequency. If the availability heuristic is applied, then such factors will affect the perceived frequency of classes and the subjective probability of events. Consequently, the use of the availability heuristic leads to systematic biases" (1973, p. 209).

As one example of how application of the availability heuristic can lead to an erroneous judgment, Tversky and Kahneman report the following experiment. Subjects were asked to estimate the number of different r-member committees that can be formed from a group of 10

people. The estimates tended to decrease with increasing r for values of r between 2 and 8. In particular, subjects typically judged it to be possible to form many more committees of size 2 than of size 8, when in fact the same number is possible in both cases. (Similar results were obtained when subjects were asked to estimate the number of different patterns of r stops that a bus could make while traversing a route with 10 stations between start and finish.) The explanation for this result, according to Tversky and Kahneman, lies in the fact that committees of two members are more readily imagined than those of eight, and, consequently, appear to be more numerous.

The major difference between the heuristic principles of representativeness and availability, Kahneman and Tversky suggest, is in the nature of the judgments on which the subjective probability estimates are based. "According to the representativeness heuristic, one evaluates subjective probability by the degree of correspondence between the sample and the population, or between an occurrence and a model. This heuristic, therefore, emphasizes the generic features, or the connotation, of the event. According to the availability heuristic, on the other hand, subjective probability is evaluated by the difficulty of retrieval and construction of instances. It focuses, therefore, on the particular instances, or the denotation, of the event. Thus, the representativeness heuristic is more likely to be employed when events are characterized in terms of their general properties; whereas, the availability heuristic is more likely to be employed when events are more naturally thought of in terms of specific occurrences" (Kahneman & Tversky, 1972, p. 452). A feature common to both heuristics is their reliance on mental effort as an indicant of subjective probability. "It is certainly harder to imagine an uncertain process yielding a nonrepresentative outcome than to imagine the same process yielding a highly representative outcome. Similarly, the less available the instances of an event, the harder it is to retrieve and construct them" (ibid, p. 452).

8.5.3 A Methodological Consideration

There is a methodological consideration relating to some of the findings of judgmental biases that deserves more attention than it has received. This has to do with the possible role of language ambiguities. We have already alluded more than once to the well known fact that the meaning of language is conditioned by the situation in which it occurs. To borrow an example from Dreyfus (1961), "a phrase like 'stay near me' can mean anything from 'press up against me' to 'stand one mile away,' depending upon whether it is addressed to a child in a crowd or a fellow astronaut exploring the moon" (p. 20). Although it seems unlikely that many of the results that have been mentioned above can be attributed to the imprecision of language, the possibility that some of them

may be based, at least in part, on this factor should not be overlooked. The finding that the estimated variability of a set of data tends to decrease as the mean increases was mentioned in a preceding section as one possible case in point. Tversky and Kahneman's finding that people judge it to be possible to form a larger number of different 2-man committees than 8-man committees from a pool of 10 men may be another. There is a way of defining "different" (e.g., "having no people in common") such that the judgment would be valid, and before one can take the results as evidence of faulty intuitions concerning combinatorics, one must be certain that none of the subjects is using such a definition. Our guess is that language ambiguities will not go far toward explaining the results obtained by Tversky and Kahneman, but it seems conceivable that they may have played some role, and some further research might be directed toward determining the extent of that role.

8.5.4 Training and Intuitive Probability Theory

We have reviewed these results at some length because this general line of research strikes us as being not only exceptionally interesting from a theoretical point of view, but of considerable practical significance. To the extent that the heuristics that have been identified are representative of the ways in which people generally make judgments of likelihood, it is clearly important to determine those conditions under which they lead to erroneous judgments and those under which they do not. Tversky and Kahneman have demonstrated that there are at least some situations in which judgments, that are presumably based on identifiable heuristics, err in systematic ways. This does not, of course, establish that these heuristics are, on balance, bad, as they are careful to point out. What one would like to know is the relative frequency with which they lead to erroneous decisions in practical real-life situations. From the point of view of the training of decision makers the question is how to foster the use of such heuristics in situations in which they are most likely to be effective, while discouraging their use in situations in which they are likely to lead to erroneous judgments. Perhaps at least a small step in that direction would be to make decision makers explicitly aware of the nature of the heuristics that tend to be used in estimating probabilities, and of the types of erroneous decisions to which they can sometimes lead.

8.6 Bayesian Inference

Undoubtedly the most widely advocated formal approach to the application of incoming data to the evaluation of hypotheses is the "Bayesian" approach. Because it has attracted so much attention and has been the focus of so much research, we will consider it in some detail.

8.6.1 Bayes Rule

It is necessary to begin with a set of mutually exclusive and exhaustive hypotheses, H_i, concerning the state of the world. To each of these hypotheses one must assign a probability, $p(H_i)$, that that hypothesis is true. Because these hypotheses are, by definition, mutually exclusive and exhaustive, it follows that the a priori probabilities sum to one, i.e.,

$$\sum_i p(H_i) = 1. \tag{1}$$

Inasmuch as the hypotheses that one is considering are likely to have different implications concerning what might be observed under specified conditions, it seems intuitively reasonable that one should be able to increase one's degree of certainty concerning the truth or falsity of any given hypothesis by making appropriate observations. For example, if H_i implies D, and if D is observed, then the credibility of H_i might reasonably be expected to be increased. (The truth of H_i is not <u>proved</u> by such an observation, of course, inasmuch as it does not <u>follow</u> from the fact that H_i implies D that D implies H_i; as was pointed out in Section 8.3, inferring the truth of H_i from the observation of D would involve the logical fallacy known as "asserting the consequent.") If both H_i and H_j could lead to D, but the likelihood of D given H_i is greater than its likelihood given H_j, then our intuitive notions about evidence suggest that the observation of D should increase our confidence in H_i somethat more than our confidence in H_j. These notions were expressed formally by the 18th Century British minister, Thomas Bayes, in the so-called "inverse probability theorem"--a theorem or rule that has been the subject of much debate.

Bayes rule expresses $p(H_i|D)$, the probability that H_i is true given the observation, or datum, D, as a function of $p(D|H_i)$, the probability that D will be observed given H_i is true, and $p(H_i)$, the probability that H_i is true as determined <u>prior</u> to the observation of D. The probability of an observation given a hypothesis, $p(D|H)$ is usually referred to as a <u>conditional</u> probability; the

probability of a hypothesis given an observation, $p(H|D)$, is usually called a posterior probability. Bayes rule defines a procedure for using the fact that D has been observed, to adjust one's estimate of the probability that H_i is true. The rule may be written as

$$p(H_i|D) = \frac{p(D|H_i)\ p(H_i)}{\sum_{j=1}^{n} p(D|H_j)\ p(H_j)}, \tag{2}$$

where n is the total number of hypotheses in the set. Because $\sum_j p(D|H_j)\ p(H_j) = P(D)$, equation (2) may be simplified to

$$p(H_i|D) = \frac{p(D|H_i)\ p(H_i)}{p(D)}. \tag{3}$$

When a sequence of observations is made, the rule is applied recursively, and the value of $p(H_i|D)$ that is computed as the result of one observation becomes the $p(H_i)$ for the following computation. That is to say, the posterior probabilities resulting from one observation become the prior probabilities for the next one. Thus, equation (3) may be written more appropriately as:

$$p_n(H_i|D) = \frac{p(D|H_i)\ p_{n-1}(H_i|D)}{p_{n-1}(D)}, \tag{4}$$

where $p_n(H_i|D)$ represents $p(H_i|D)$ after the n^{th} observation, and $p_0(H_i|D)$ or, more appropriately, $p_0(H_i)$, is understood to be the probability of H_i before any observations are made. We will follow the convention of using subscripts only when they are essential for clarity.

Bayes rule states, in effect, that if the prior probability of a hypothesis being true, $p(H_i)$ and the probability of observing a particular datum given that hypothesis is true, $p(D|H_i)$ are known for all i, then the probability that the hypothesis is true given

that the datum has been observed, $p(H_i|D)$, can be calculated in a straightforward way. In many decision situations, $p(H_i)$ and $p(D|H_i)$ are not known, and cannot be determined objectively; therefore, they must be estimated. The significance of the rule stems from the assumption, for which there is some evidence that will be considered later, that people are better at estimating conditional probabilities, $p(D|H)$, than at estimating posterior probabilities, $p(H|D)$. Obviously, if they were invariably very good at estimating $p(H|D)$ there would be no need to make use of Bayes rule to calculate this value; it would suffice to have the decision maker estimate it directly.

8.6.2 Likelihood Ratio

In order to make use of Bayes rule it is not necessary to require that an individual estimate probabilities explicitly. An alternative procedure is to have him judge the ratios of pairs of conditional probabilities. Such ratios are referred to as likelihood ratios. The likelihood ratio of D given H_1 relative to D given H_2 may be expressed as follows:

$$L_{1,2} = \frac{p(D|H_1)}{p(D|H_2)} . \qquad (4)$$

The attractiveness of likelihood ratio stems from the fact that people often find it easier to make the implied judgment than to estimate conditional probabilities directly. The type of judgment that is required in this case is of the sort "Event D is X times as likely if H_1 is true than if H_2 is true." Neither of the conditional probabilities need be specified explicitly. A disadvantage associated with its use is the fact that a great many more judgments are required with respect to each observation.

8.6.3 Other Methods for Obtaining Probability Estimates

Other methods have been used to obtain probability estimates without having the subject explicitly produce numerical values. For chips-in-urn problems, for example, Peterson and Phillips (1966) have had subjects adjust markers on a scaled 0-to-1 continuum so that each interval is equally likely to contain the true proportion of chips of a specified color. Organist (1964) developed a simple answer chart which forced a subject to make his distribution of probabilities over the possible hypotheses sum to one and also specified what his payoff would be for each

hypothesis if it were correct, given the probability that he attached to it. Shuford (1967) describes a computer-controlled system which presents a subject with a set of hypotheses and allows him to specify probabilities by adjusting the lengths of lines associated with the hypotheses by pointing at them with a light pen. When one line is lengthened or shortened, compensatory adjustments are automatically made in the remaining lines so that the probabilities always sum to one. This system also provides the user with information concerning the implications of his probability assignments vis-a-vis his payoff, given the truth of any particular hypothesis.

8.6.4 Diagnosticity of Data

Intuition suggests that the more disparate the implications of two hypotheses, the more informative data should b concerning which of the hypotheses is likely to be true. In a Bay sian context the informativeness, or "diagnosticity," of data is defined in terms of the likelihood ratio. Specifically, the magnitude of a likelihood ratio is said to represent the diagnosticity of a datum with respect to the two particular hypotheses involved. The more the ratio **differs** from 1:1, in either direction, the more informative the datum is with respect to which of the hypotheses under consideration is correct, and the more the distribution of probabilities over these hypotheses will change as a consequence.

8.6.5 Odds

The ratio of two posterior probabilities is referred to as the posterior "odds" with respect to the associated hypotheses. The posterior odds of H_1 with respect to H_2 may be expressed as

$$\frac{p_n(H_1|D)}{p_n(H_2|D)} = \left[\frac{p(D|H_1)\ p_{n-1}(H_1|D)}{p(D)}\right] \Big/ \left[\frac{p(D|H_2)\ p_{n-1}(H_2|D)}{p(D)}\right] \quad (6)$$

or, equivalently, as

$$\frac{p_n(H_1|D)}{p_n(H_2|D)} = \frac{p(D|H_1)}{p(D|H_2)} \cdot \frac{p_{n-1}(H_1|D)}{p_{n-1}(H_2|D)}, \quad (7)$$

which is to say that the posterior odds is simply the prior odds multiplied by the likelihood ratio. Letting $\Omega_{n;i,j}$ represent the odds of H_i with respect to H_j after the n^{th} observation, we may express the relationship as follows;

$$\Omega_{n;i,j} = L_{i,j}\Omega_{n-1;i,j} \, . \tag{8}$$

Obviously

$$\Omega_{j,i} = \Omega_{i,j}^{-1} \tag{9}$$

and

$$L_{j,i} = L_{i,j}^{-1}. \tag{10}$$

Often it is clear from the context which of the two terms of either an odds ratio or a likelihood ratio is to be the numerator and which the denominator, so the subscripts are omitted and the expression is written more simply as

$$\Omega_n = L\Omega_{n-1}. \tag{11}$$

It is essential, however, that the same hypothesis, whether H_i or H_j, be represented in the same position (numerator or denominator) in both ratios.

8.6.6 Applications of Bayes rule in The Two-Hypothesis Case.

To summarize what has been said so far, Bayes rule represents a procedure for evaluating hypotheses in situations that have the following characteristics: (a) the possible states of the world can be explicitly represented by an exhaustive, and mutually exclusive set of possibilities; (b) discrete observations may be made in an effort to find more information about the actual state of the world; and (c) for the data obtained from each observation, it must be reasonable to assign a number that represents the probability that those data would have been obtained, given the truth of any specific one of the hypothesized states of the world. In order to get an appreciation of how Bayes rule extracts information from data, it will be helpful to consider some concrete examples of decision tasks to which the rule might be applied. We will focus first on the simple case in which the hypothesis set contains only two alternatives.

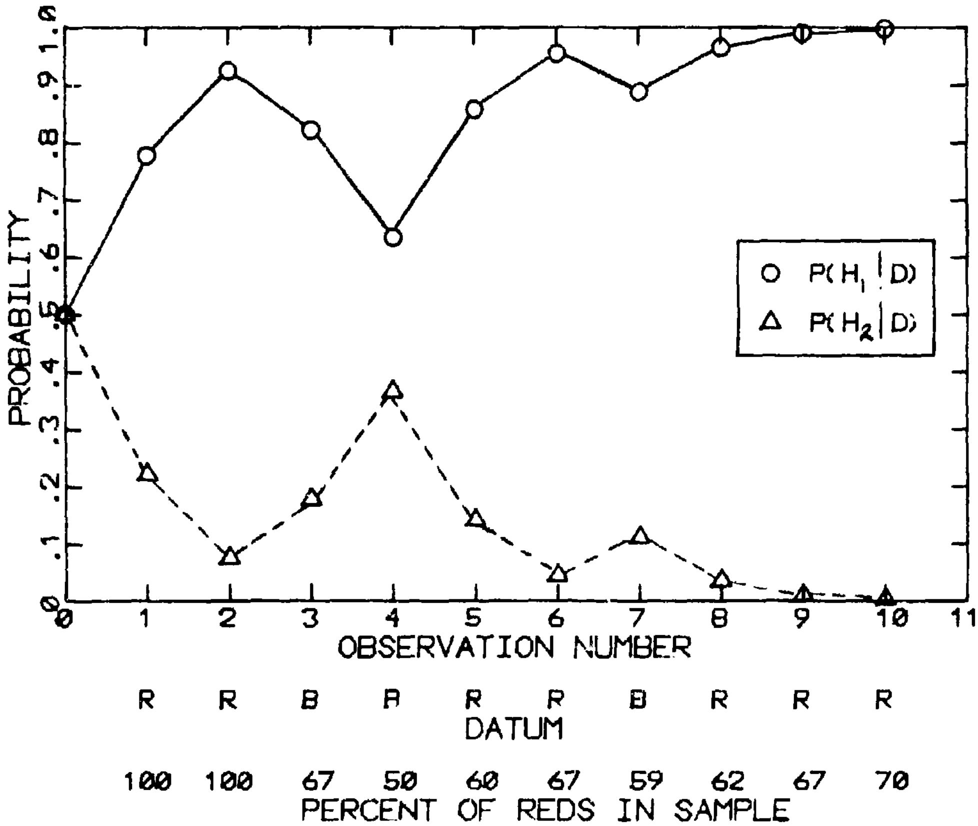
Imagine an urn containing red and black chips. Suppose two hypotheses, H_1 and H_2, are stated, one and only one of which is true, concerning what proportion of the chips in the urn are red. The task is to decide which of these hypotheses is the true one. Data are obtained by sampling chips one at a time, replacing each chip after it is examined. Assume that the chips are thoroughly mixed before each observation and that the probability of drawing a red chip on a trial is exactly R/R+B, where R is the number of red chips, and B the number of black chips, in the urn.

Suppose the first hypothesis, H_1, is that 70% of the chips are red, and that the second hypothesis, H_2, is that 20% of the chips are red. Suppose further that the prior probabilities are equal, that is, $p_0(H_1) = p_0(H_2) = .5$. Figure 1 shows how $p(H_1|D)$ and $p(H_2|D)$ change as a result of applying Bayes rule to the data obtained in the following ten successive observations: RRBBRRBRRR. Figures 2 and 3 show the odds, and the uncertainty, in the information theoretic sense of the word, change from observation to observation. Uncertainty is, of course, a monotone but nonlinear function of the difference between the probabilities associated with the two hypotheses.

Note that the effect of an observation is not necessarily to decrease the amount of uncertainty concerning which hypothesis is true. If the distribution of $p_0(H_i)$ favors the incorrect hypothesis, uncertainty is very likely to increase as a result of observing data before it decreases. Even if the distribution of $p_0(H_i)$ favors the correct hypothesis, or weights both hypotheses equally, uncertainty may increase on individual trials. In this case, however, it will decrease <u>on the average</u>, assuming unbiased sampling.

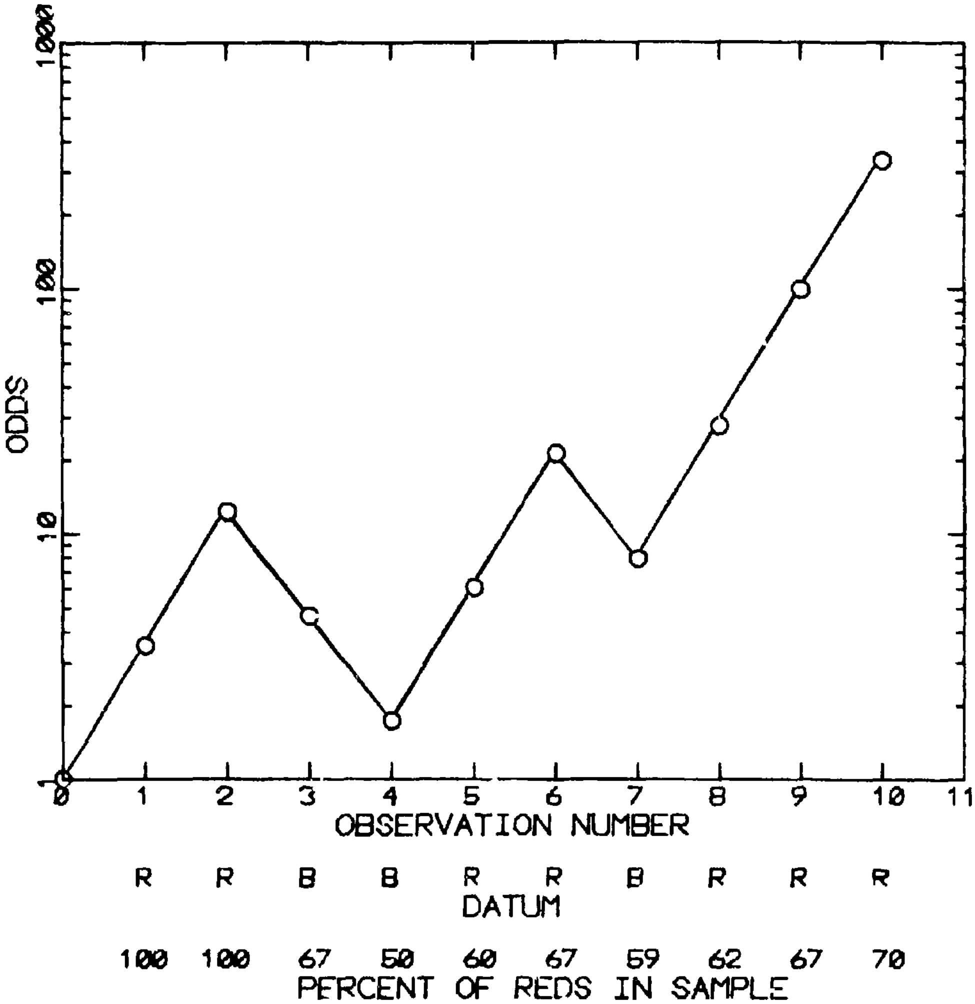
Another interesting and perhaps counterintuitive observation concerning figure 1 is the very large effect that the one or two initial observations can have. In our example, the initial drawing of two successive reds had the result of making one of the (initially equally likely) hypotheses over twelve times more likely than the other.

Intuitively, one would expect that the degree of confidence that one should have that the proportion of reds and blacks in one's sample reflects the true proportion in the population should depend on the sample size. That the application of Bayes theorem does not violate this intuition may be seen by comparing the probability distribution after the third observation and after the sixth observation (figure 1). In both cases, red chips have comprised 67 percent



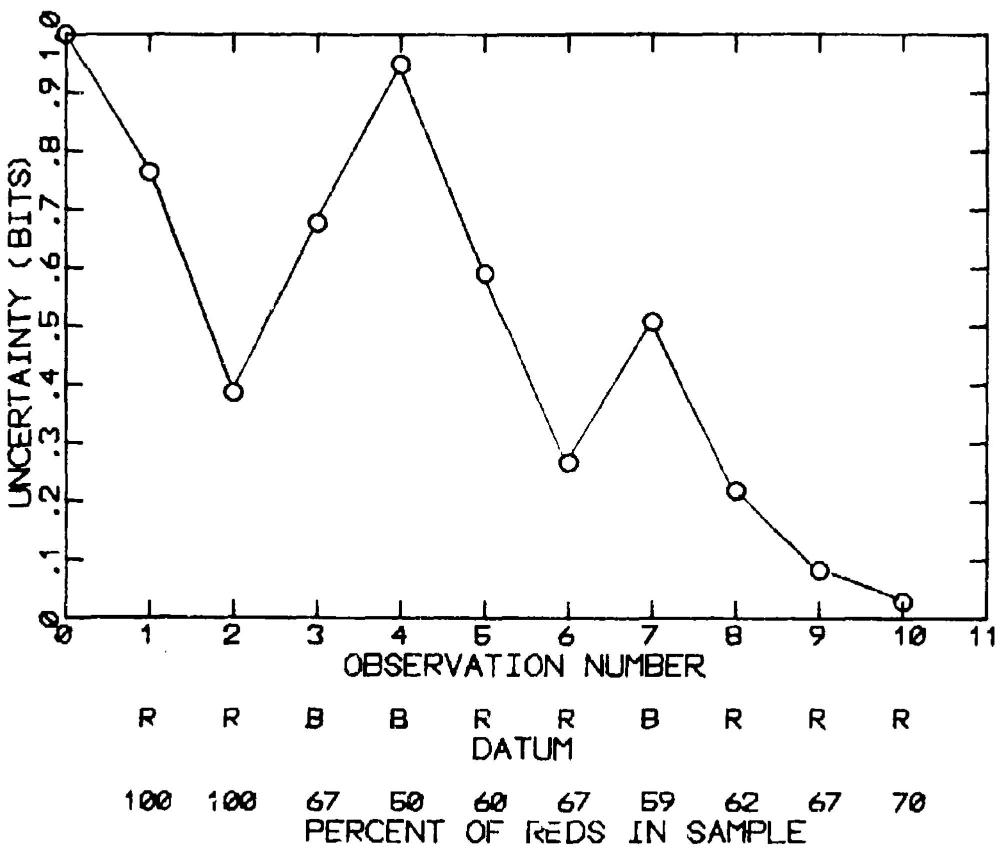
H_1: 70% Red; H_2: 20% Red;
$p_0(H_1) = p_0(H_2) = .5$

Figure 1. Changes in posterior probabilities, $p(H_1|D)$ and $p(H_2|D)$ as a result of the indicated observations of Red and Black chips



H_1: 70% Red; H_2: 20% Red;
$p_0(H_1) = p_0(H_2) = .5$

Figure 2. Changes in odds, $\Omega_{1,\ 2}$ as a result of the indicated observations of Red and Black chips



H_1: 70% Red; H_2: 20% Red;
$p_0(H_1) = p_0(H_2) = .5$

Figure 3. Changes in uncertainty concerning hypotheses as a result of the indicated observations of Red and Black chips

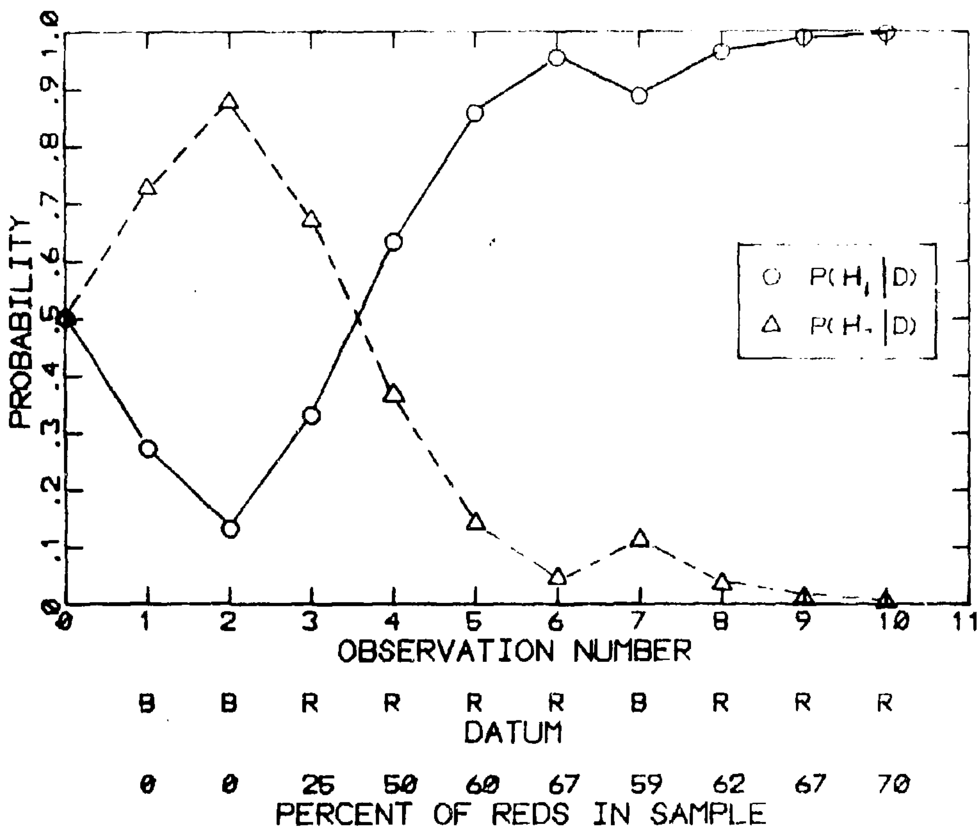
of the sample; however, the uncertainty is less following the sixth observation than following the third, reflecting the fact that the sample size was larger in the former case.

Figure 4 shows how the probabilities change over the course of ten observations in which reds and blacks occur with the same frequency but in a different order. In particular, the first two observations in this case produced blacks, and the second two, reds. Observations 5 through 10 are assumed to be the same as in the original example. Note that by the end of the fourth trial, the proportion of red and black draws was the same in both examples; consequently, the probability distributions are the same at this point and thereafter. This illustrates the fact that the Bayesian calculation of $p(H_i|D)$ is path-independent, in the sense that the effect of an observation is strictly dependent on the current value of $p(H_i)$, and independent of the particular sequence of observations on which that value is based. The calculation is also independent of the number of observations on which the current value of $p(H_i)$ is based. Note that this point is different from the one made above concerning the effect of sample size on uncertainty. The point that was made above was that the probability that a given proportion of reds in one's sample accurately reflects the proportion in the population increases with sample size. The point here is that the effect that an observation will have is independent of how $p(H_i)$ got to be whatever it is.

Figures 5, 6 and 7 illustrate the effects of setting the initial values of $p(H_1)$ and $p(H_2)$ to something other than .5. The sequence of draws is identical to that in figure 1, and consistent with what might be expected if the true hypothesis were H_1. In each figure, one curve shows the effect of these observations given that $p_0(H_1) = .8$; another shows the effect given that $p_0(H_1) = .2$, and the third represents $p_0(H_1) = .5$. The main thing to notice is that the effect of an initial incorrect bias is largely nulled out by relatively few observations. This point is frequently made by proponents of Bayesian information-processing systems in response to the observation that a priori probabilities are sometimes difficult to assign on anything other than an arbitrary basis. A fact that usually is not pointed out is illustrated in figure 7: changing the distribution of a priori probabilities shifts the function relating log odds to data by a constant.

8.6.7 Expected Effects of Observations on Hypotheses

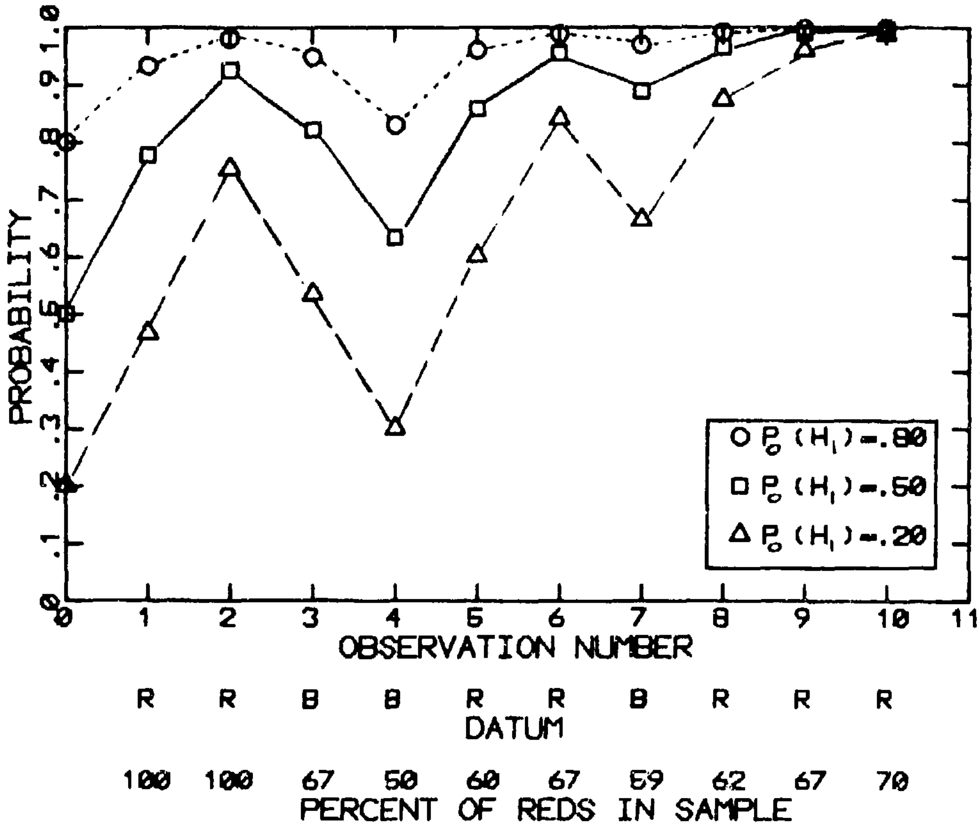
In the foregoing examples of the application of Bayes rule, we have considered how probabilities may change as a result of a sequence of specific observations. It has been apparent from these examples that the effect of an observation sometimes is to increase the probability associated with the true hypothesis and sometimes to decrease it. On the average, however, we expect the probability



H_1: 70% Red; H_2: 20% Red;
$p_0(H_1) = p_0(H_2) = .5$

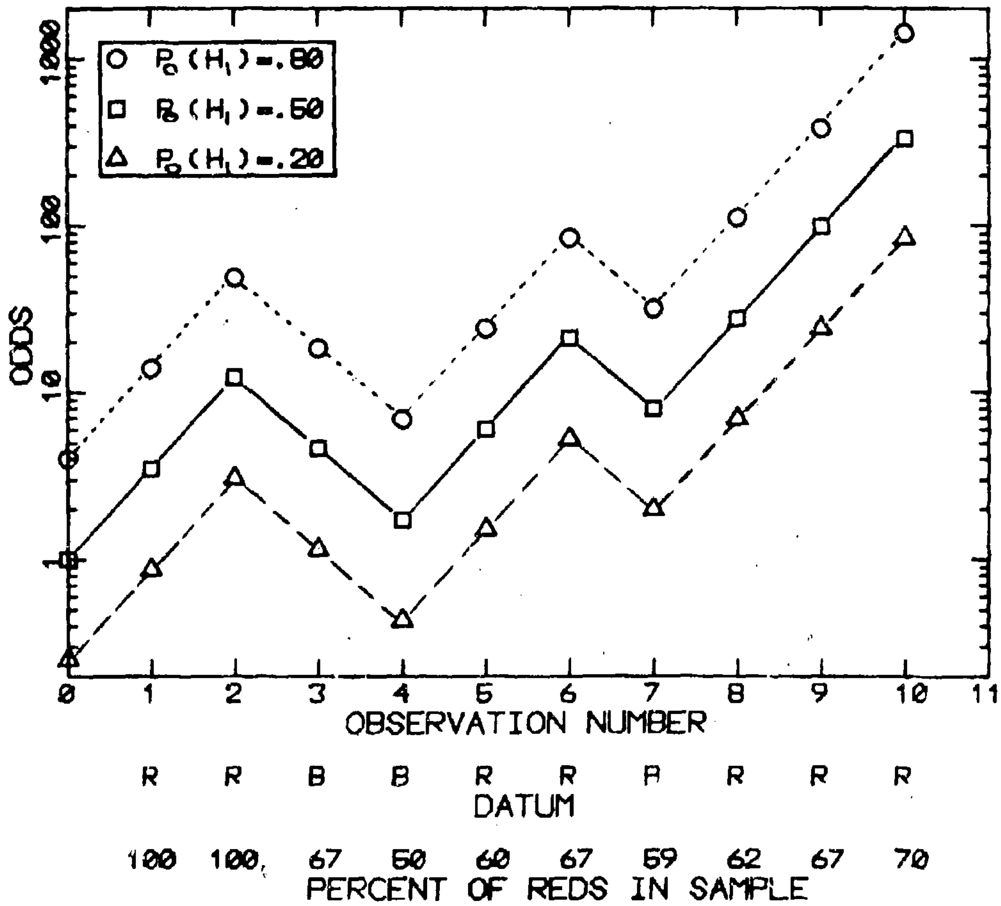
Figure 4. Changes in posterior probabilities, $p(H_1|D)$ and $p(H_2|D)$ as a result of the indicated observations of Red and Black chips

(Note that the results of the observations are the same as in figure 1 except that the first four produce a different ordering of Reds and Blacks.)



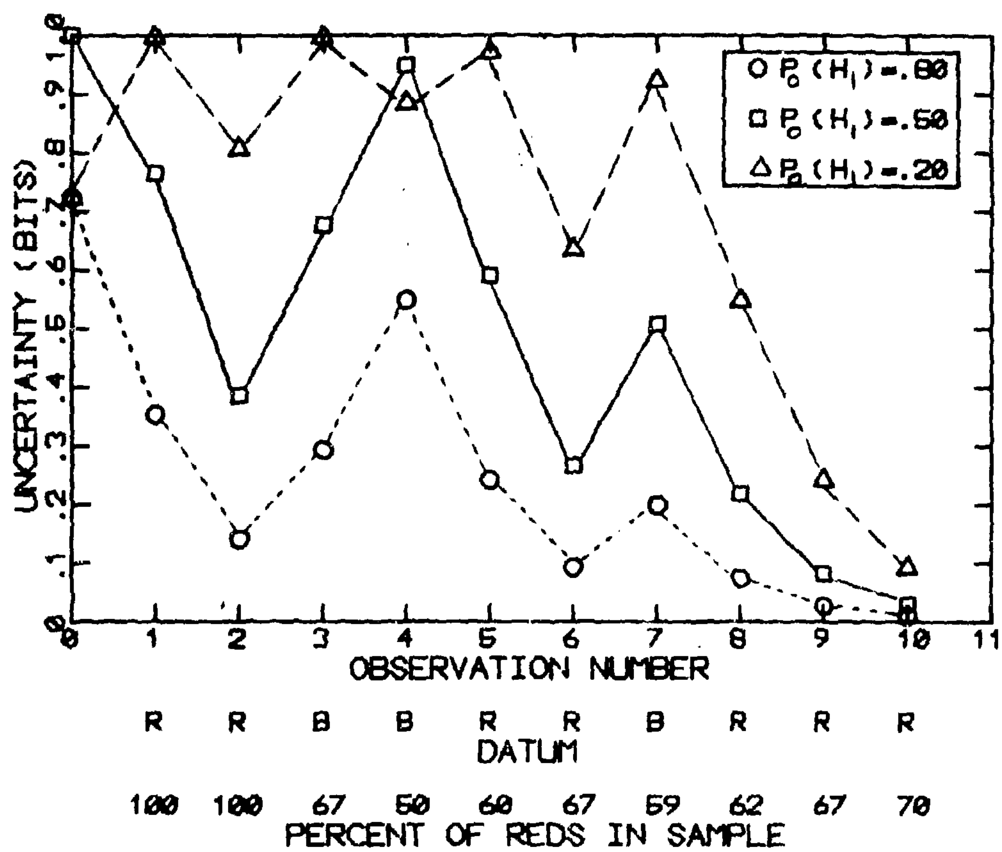
H_1: 70% Red; H_2: 20% Red

Figure 5. Effects of indicated observations on $p(H_1|D)$ for different values of $p_0(H_1)$



H_1: 70% Red; H_2: 20% Red

Figure 6. Effects of indicated observations on odds, $\Omega_{1,2}$ for different initial values of $p(H_1)$



H_1: 70% Red; H_2: 20% Red

Figure 7. Effects of indicated observations on uncertainty for different initial values of $p(H_1)$

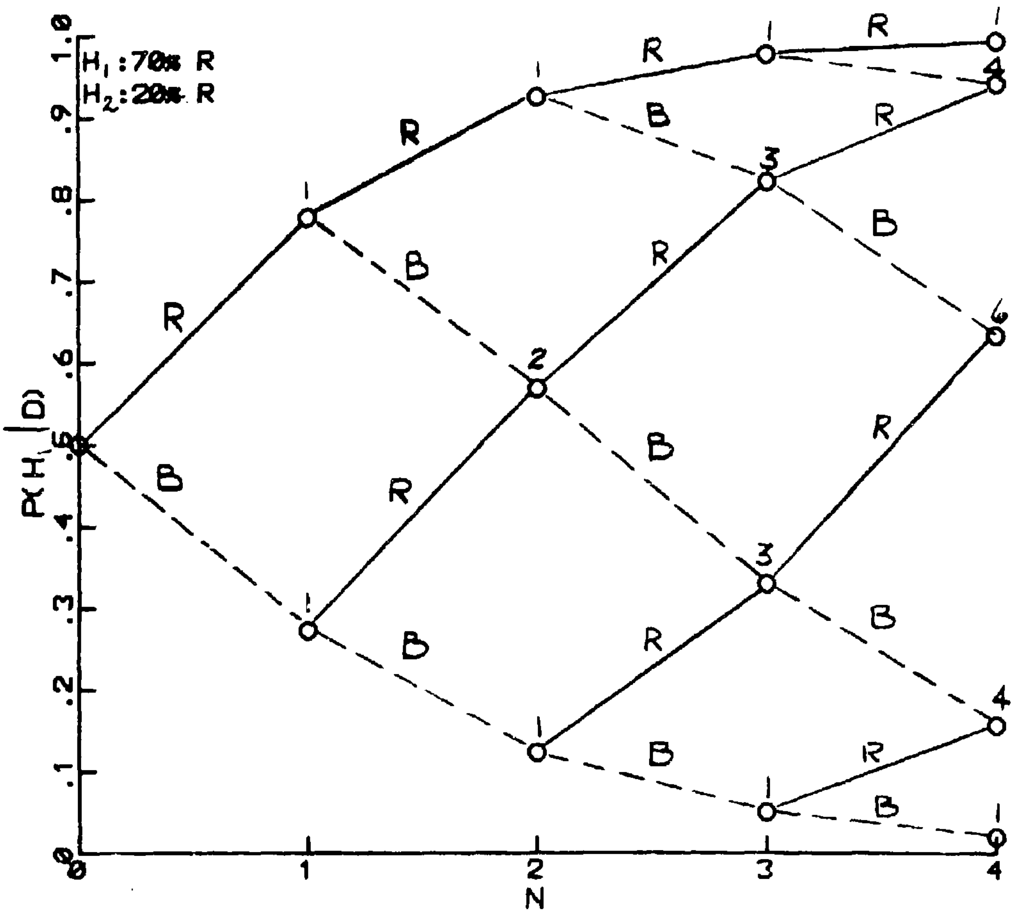
associated with the true hypothesis to increase with each observation, and that associated with a false hypothesis to decrease, assuming an unbiased sampling of the data. We turn now to a consideration of how $p(H_i|D)$ can be expected to change, on the average, as a result of applying Bayes rule, if chips are drawn from an urn that contains reds and blacks in the proportion indicated by a specified hypothesis, say H_j.

It will facilitate the discussion to begin by considering all possible outcomes of a specific experiment. Figure 8 shows all possible effects of four observations on $p(H_1|D)$, in the case of our example of H_1: 70% R, H_2: 20% R, and $p(H_i) = .5$. Each node in the graph represents one possible value of $p(H_1|D)$ after the number of observations indicated on the abscissa; no values are possible other than those represented by nodes. (By rotating the graph in figure 8 about a horizontal axis passing through the .5 point on the ordinate, one would produce the graph of $p(H_2|D)$; which is to say, each of the points in the graph of $p(H_2|D)$ is the complement of a point in the graph of $p(H_1|D)$.) In general, after N observations, $p(H_i|D)$ will have one of N+1 possible values. After two observations, for example, $p(H_1|D)$ will have one of the three values .925, .568, or .123. The number above each node indicates the number of ways to arrive at that node. There are three ways, for example, to arrive at the node at $p(H_1|D) = .821$, $N = 3$: RRB, RBR and BRR. The set of numbers associated with a given value of N will be recognized as the coefficients of the terms of the expansion of $(a+b)^N$, the so-called "binomial coefficients." In our application, each of these coefficients, which may be written as $\binom{N}{m}$, represents the number of ways that N events can be composed of m events of one type and N-m of another. The events of interest in our case are draws of chips from an urn, and the two types are draws of red and black chips, respectively. The sum of these coefficients for given N,

$$\sum_{m=0}^{N} \binom{N}{m} = 2^N, \tag{12}$$

represents the number of uniquely ordered sequences of reds and blacks that can result from N draws. Inasmuch as the effect of applying Bayes rule to a sequence of data is insensitive to the order in which the data are considered, it is convenient to think of all sequences having the same combination of reds and blacks as the same outcome, irrespective of the order in which the reds and blacks have occurred. Thus, the effective number of possible outcomes of N draws is N+1 rather than 2^N.

Figure 8 shows the graph of possible outcomes for our hypothetical experiment as they pertain to $p(H_1|D)$. By the



H_1: 70% Red; H_2: 20% Red;
$p_0(H_i) = .5$

Figure 8. All possible values of $p(H_1|D)$ after N observations

algebra of expectation, the expected, or mean value of $p(H_i|D)$ after N draws is the weighted sum of all possible values, each value being weighted by its probability of occurrence. In terms of figure 8, the expected value of $p(H_i|D)$ after N draws may be found by multiplying the product of the value of each node above N on the abscissa by the probability of arriving at that node, and summing over these products.

Suppose that the probability that an observation will yield a red chip (solid lines) is q, and the probability that it will yield a black one (dotted lines) is 1-q. The probability of arriving at a given node in the graph, via a particular path, is the product of the probabilities associated with the links in that path. The probability of arriving at a given node, irrespective of the path, is the sum of the probabilities associated with all possible paths to that node. But every path leading to a common node has exactly the same probability of being traversed, because each is composed of the same combination of R and B links. So, the easy way to calculate the probability of arriving at a node is to take the product of the probability of traversing any specific path to that node and the number of paths leading to that node. Figure 9 shows expressions for these probabilities for each of the nodes in our sample graph. In general, the probability of arriving at a given node via a specific path composed of m R links and N-m B links, is given by

$$q^m(1-q)^{N-m}$$

and the probability of arriving at a node via any such path by

$$\pi_{N,m} = \binom{N}{m} q^m (1-q)^{N-m} . \qquad (13)$$

The expected value of $p(H_i|D)$ after N observations, then, is given by

$$E\left[p_N(H_i|D)\right] = \sum_{m=0}^{N} \pi_{N,m} P_{N,m} , \qquad (14)$$

where $P_{N,m}$ represents the posterior probability of H_i after N observations, m of which have yielded red chips.

The following iterative formula can be used to compute π :

$$\pi_{N,m+1} = \frac{(N-m)q}{(m+1)(1-q)} \pi_{N,m} \quad , \tag{15}$$

where

$$\pi_{N,0} = (1-q)^N, \tag{16}$$

and computation can be simplified by taking logarithms:

$$\log \pi_{N,m+1} = \log x + \log \pi_{N,m} \quad , \tag{17}$$

where

$$x = \frac{(N-m)(q)}{(m+1)(1-q)} \quad , \tag{18}$$

and

$$\log \pi_{N,0} = N \log (1-q) \cdot \tag{19}$$

The value of q in equation (13) depends, of course, on which of the hypotheses under consideration happens to be true. The expectation can be computed, however, for each of the possibilities. The general expression may be written as follows:

$$E\left[p_N(H_i|D) \middle| H_j \text{ is true}\right] = \sum_{m=0}^{N} \binom{N}{m} p(R|H_j)^m p(B|H_j)^{N-m} P_{N,m} \cdot \tag{20}$$

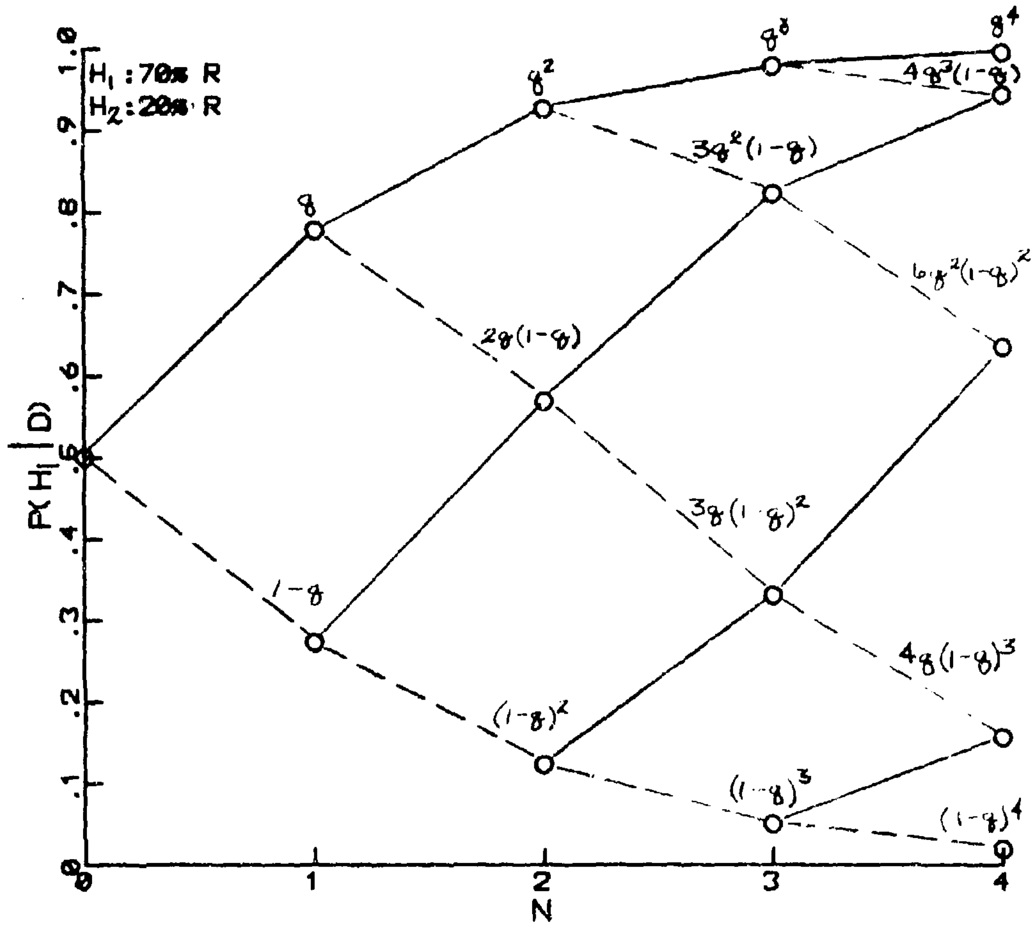


Figure 9. Graph illustrating the computation of expected values of posterior probabilities (The expression above each node represents the probability of arriving at that node, given q is the probability of drawing a red chip. The expected value of the posterior probability following N observations is the sum of the values of the nodes above N, each weighted by its "arrival" probability.)

Figure 10 shows $E[p_N(H_i|D)]$ for our example (H_1: 70% red, H_2: 20% red), given H_i is true (top curve) and given H_i is not true (bottom curve). The top curve shows the expected growth of $p(H_1|D)$ when H_1 is true and of $p(H_2|D)$ when H_2 is true. Conversely, the bottom curve represents the expected decline of $p(H_1|D)$ when H_2 is true and of $p(H_2|D)$ when H_1 is true. Thus, in the two-alternative case,

$$E\left[\, p(H_1|D) \,\middle|\, H_1 \text{ is true}\right] = E\left[\, p(H_2|D) \,\middle|\, H_2 \text{ is true}\right] . \quad (21)$$

To compute the expected uncertainty following N observations, one must compute the uncertainty associated with each of the possible outcomes of the observations, and then take a weighted average of these uncertainties, the weights being the probabilities of occurrence of the specific outcomes. The uncertainty associated with a specific outcome, say the outcome N observations yielding m red chips, is given by

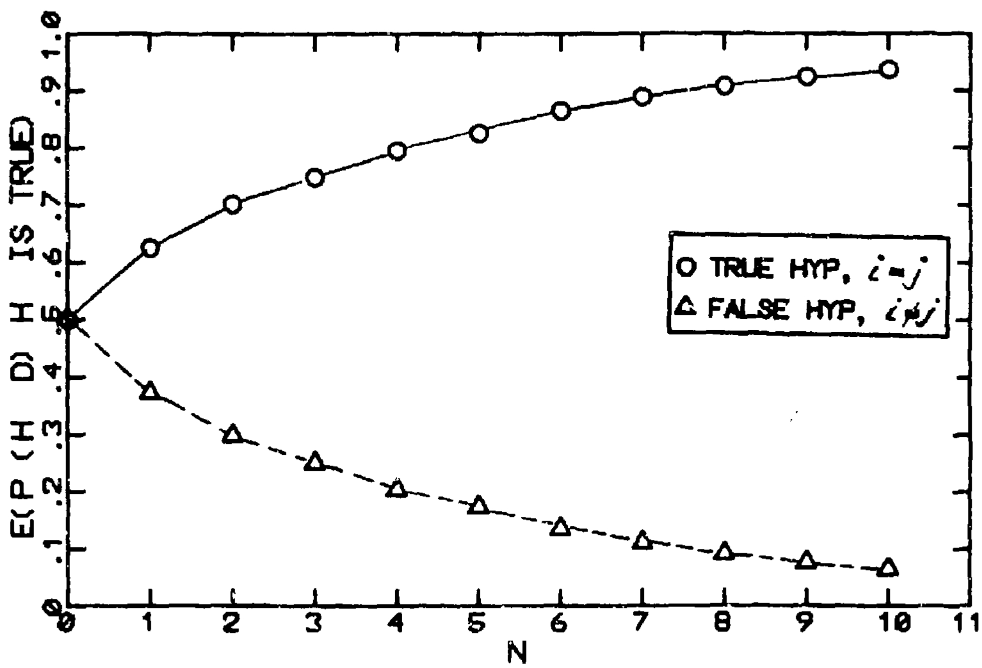
$$U_{N,m} = -\sum_{i=1}^{h} P_{i;N,m} \log_2 P_{i;N,m} , \quad (22)$$

where h is the number of hypotheses under consideration and $P_{i;N,m}$ is the probability associated with the i^{th} hypothesis after N observations yielding m red chips. The expected uncertainty after N observations, then, is obtained by weighting each $U_{N,m}$ by its probability of occurrence, and summing over all possible outcomes. Thus,

$$E(U_N) = \sum_{m=0}^{N} \pi_{N,m} U_{N,m} , \quad (23)$$

where $\pi_{N,m}$ is defined as before. Again, inasmuch as the value of q in equation (13) depends on which hypothesis is true, the general expression for $E(U_N)$ conditional upon which hypothesis is true may be written

$$E(U_N|H_j \text{ is true}) = \sum_{m=0}^{N} \binom{N}{m} p(R|H_j)^m p(B|H_j)^{N-m} U_{N,m} . \quad (24)$$



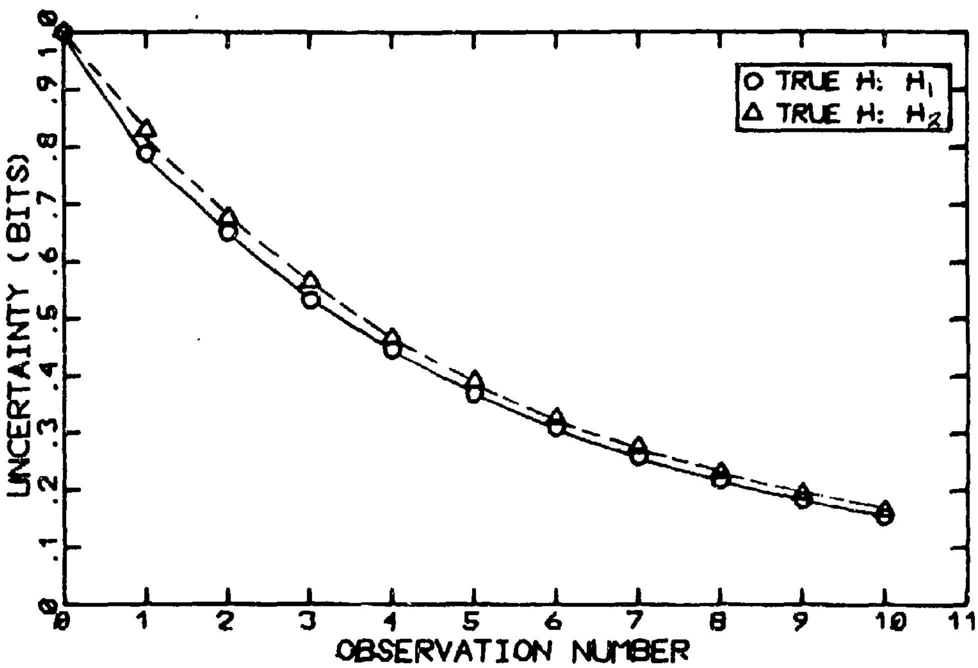
H_1: 70% Red; H_2: 20% Red;
$p_0(H_i) = .5$

Figure 10. Expected value of posterior probability H_i, given H_j is true, as a function of number of observations

The computation of expected uncertainty relates to figure 8 in the following way. Imagine that such an outcome graph were developed for each of the hypotheses under consideration, conditional upon the truth of a specified hypothesis. The value of $U_{N,m}$ is found by summing $-p \log_2 p$ across graphs for given N,m, the values of p being the values of the nodes on the graphs (equation 22). The value of $E(U_N)$ is then obtained by weighting each of these sums by the probability of arriving at node N,m, given the truth of the specified hypothesis, and summing over m (equation 23). Figure 11 shows how the expected uncertainty concerning which of the two hypotheses is true changes as a result of observations in the case of our example (H_1: 70% red, H_2: 20% red), given the truth of each hypothesis in turn.

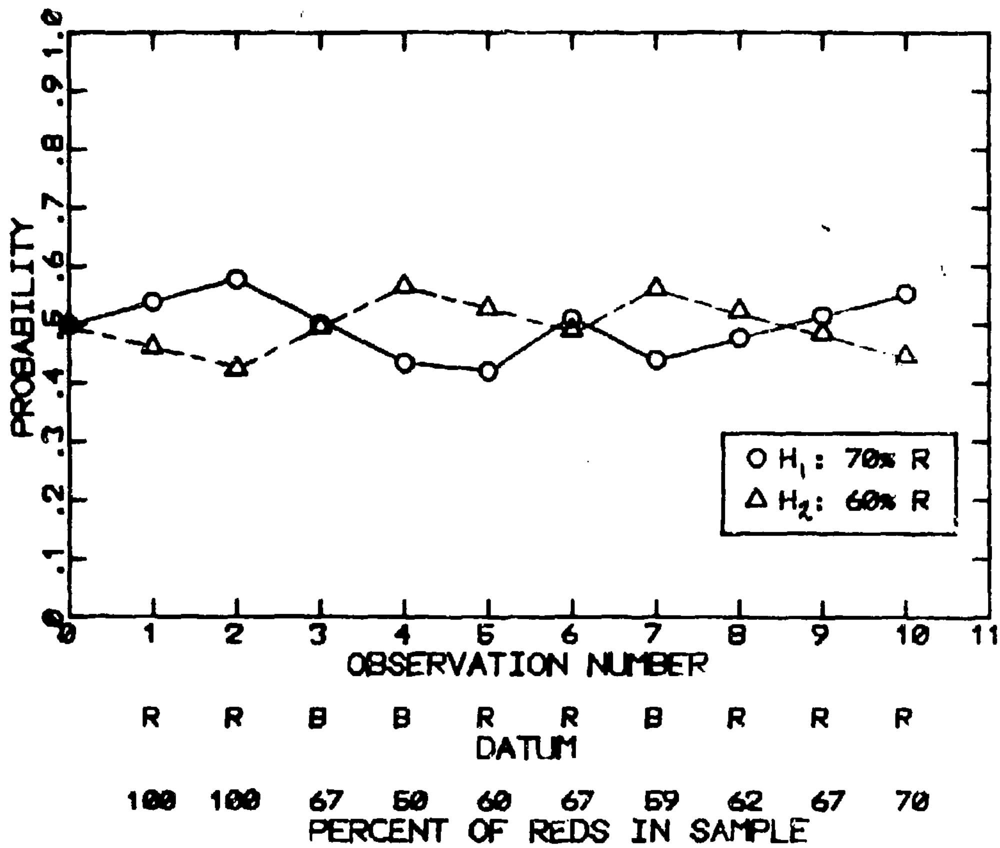
The examples that we have been considering have all converged rather quickly to a state of relatively low uncertainty. This was due to the fact that H_1 and H_2 were quite disparate. But suppose H_1 and H_2 were similar with respect to their implications for data. Suppose, for example, that we let H_1 be the same as before (that is, that 70% of the chips are red) and H_2 be the hypothesis that 60% of the chips are red. Again, setting the initial probabilities equal to .5 and assuming the same sequence of observations as indicated in figure 1, figures 12 and 13 show the effects of these observations on the distribution of probabilities over the two hypotheses, and on uncertainty. Figures 14 and 15 show the expected effects of data on posterior probabilities and uncertainty for this case. Obviously, the expected effects of observations are much smaller--the data have less diagnostic impact--when the hypotheses are similar than when they are very different. Or, to say the same thing in other words, a larger sample is needed to produce the same degree of certainty with respect to which hypothesis is true. This illustrates the intuitively compelling idea that the smaller the differences between two statistical distributions, the closer one must examine them to tell which is which. Continued sampling will eventually make the probabilities diverge and the uncertainty decrease, assuming, of course, that the sampling is random and one of the hypotheses is in fact true. Figures 16 and 17 show the expected changes in $p(H_i|D)$ and uncertainty over the course of 200 observations, given H_1: 70% red, H_2: 60% red, $p_0(H_i) = .5$. Two hundred observations would not, on the average, reduce the uncertainty in this case by the amount that ten observations would reduce it, given the more disparate hypotheses, H_1: 70% red and H_2: 20% red.

Table 2 (page 95) shows, for various combinations of H_1 and H_2, the expected posterior probability of H_i after ten observations, given that chips are sampled from an urn containing reds and blacks in the proportions specified by H_i, and assuming $p_0(H_1) = p_0(H_2) = .5$.



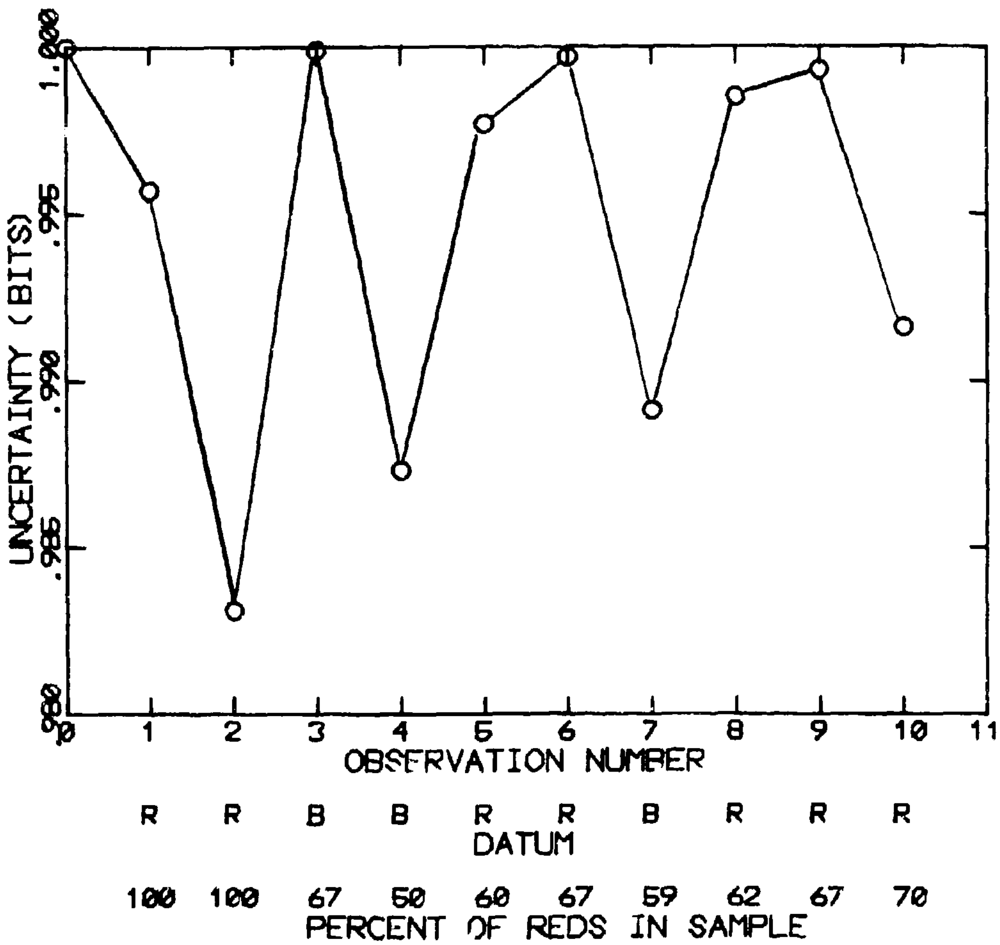
H_1: 70% Red; H_2: 20% Red;
$p_0(H_i) = .5$

Figure 11. Expected uncertainty concerning which hypothesis is true, as a function of number of observations



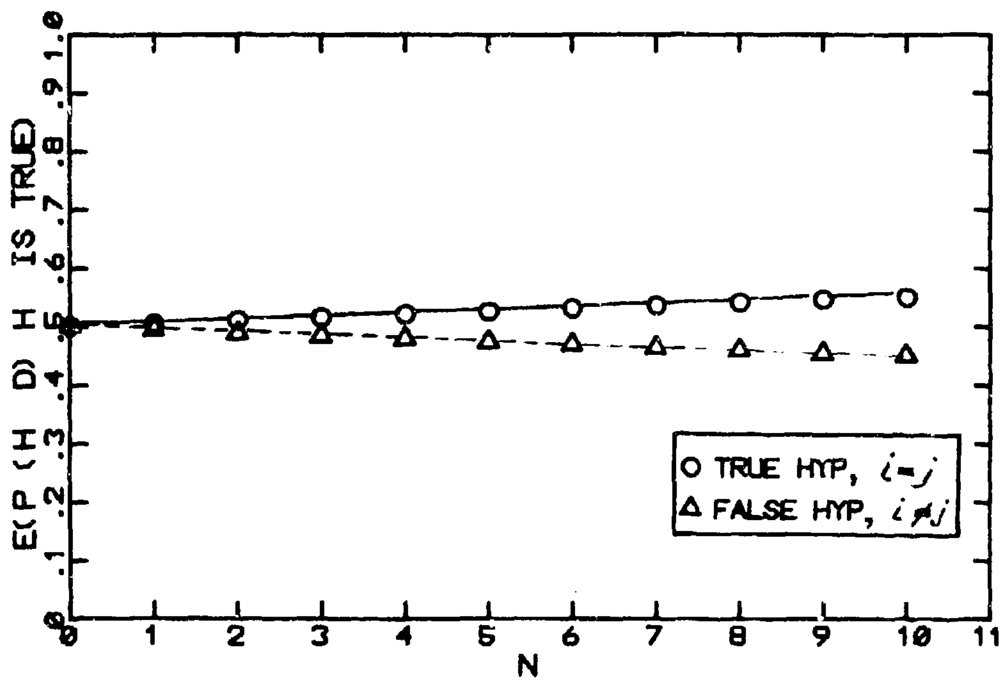
H_1: 70% Red; H_2: 60% Red;
$p_0(H_1) = p_0(H_2) = .5$

Figure 12. Changes in posterior probabilities, $p(H_1|D)$ and $p(H_2|D)$, as a result of the indicated observations of Red and Black chips



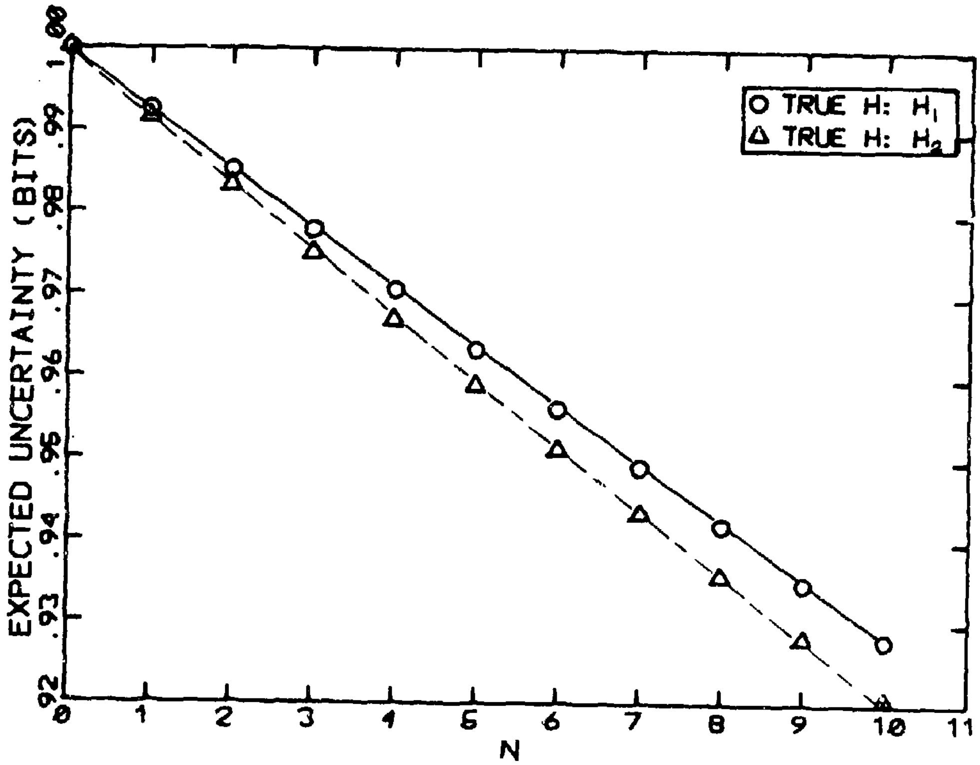
H_1: 70% Red; H_2: 60% Red;
$p_0(H_i) = .5$

Figure 13. Changes in uncertainty concerning hypotheses as a result of the indicated observations of Red and Black chips



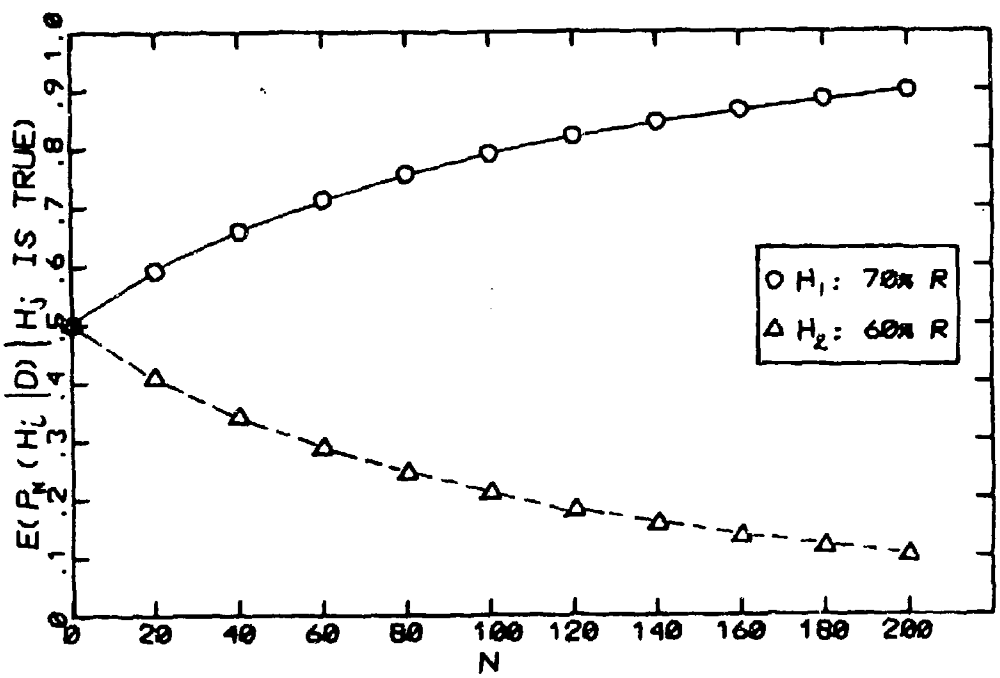
H_1: 70% Red; H_2: 60% Red;
$p_0(H_i) = .5$

Figure 14. Expected value of posterior probability of H_i, given H_j is true, as a function of number of observations



H_1: 70% Red; H_2: 60% Red;
$p_0(H_i) = .5$

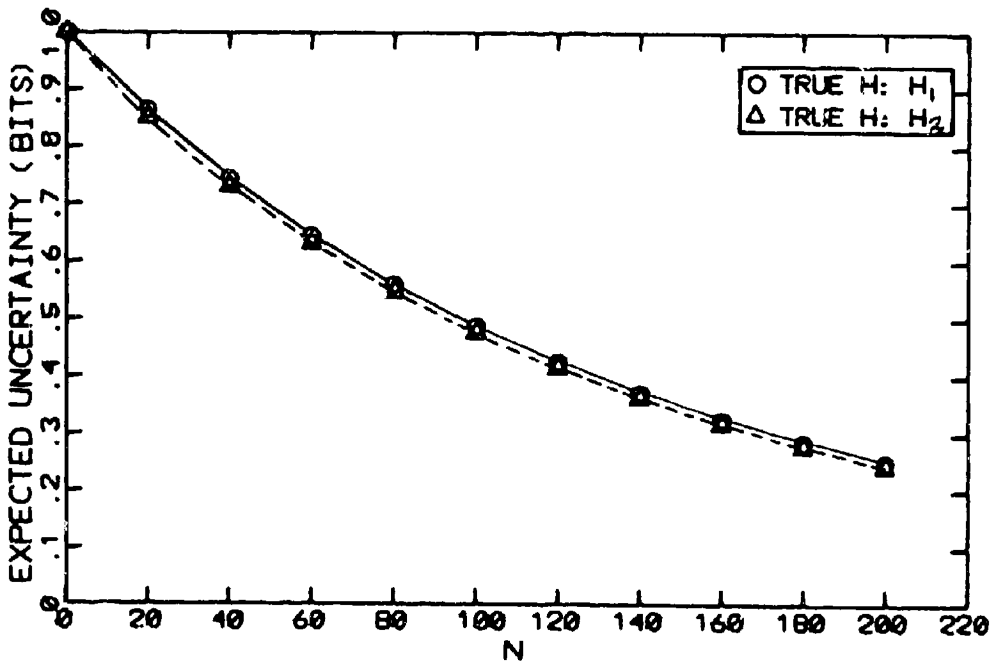
Figure 15. Expected uncertainty concerning which hypothesis is true, as a function of number of observations



H_1: 70% Red; H_2: 60% Red;

$p_0(H_i) = .5$

Figure 16. Expected value of posterior probability H_i, given H_j is true, as a function of number of observations



H_1: 70% Red; H_2: 60% Red;
$p_0(H_i) = .5$

Figure 17. Expected uncertainty concerning which hypothesis is true, as a function of number of observations

TABLE 2. SENSITIVITY OF BAYESIAN ANALYSIS.

Percentage of Reds according to H_1	Percentage of Reds according to H_2: 00	10	20	30	40	50	60	70	80	90	100
00	.50	.74	.90	.97	.99	1.00	1.00	1.00	1.00	1.00	1.00
10	.74	.50	.59	.72	.83	.90	.95	.98	.99	1.00	1.00
20	.90	.59	.50	.56	.67	.79	.87	.94	.97	.99	1.00
30	.97	.72	.56	.50	.55	.66	.77	.87	.94	.98	1.00
40	.99	.83	.67	.55	.50	.55	.65	.77	.87	.95	1.00
50	1.00	.90	.79	.66	.55	.50	.55	.66	.79	.90	1.00
60	1.00	.95	.87	.77	.65	.55	.50	.55	.67	.83	.99
70	1.00	.98	.94	.87	.77	.66	.55	.50	.56	.72	.97
80	1.00	.99	.97	.94	.87	.79	.67	.56	.50	.59	.90
90	1.00	1.00	.99	.98	.95	.90	.83	.72	.59	.50	.74
100	1.00	1.00	1.00	1.00	1.00	1.00	.99	.97	.90	.74	.50

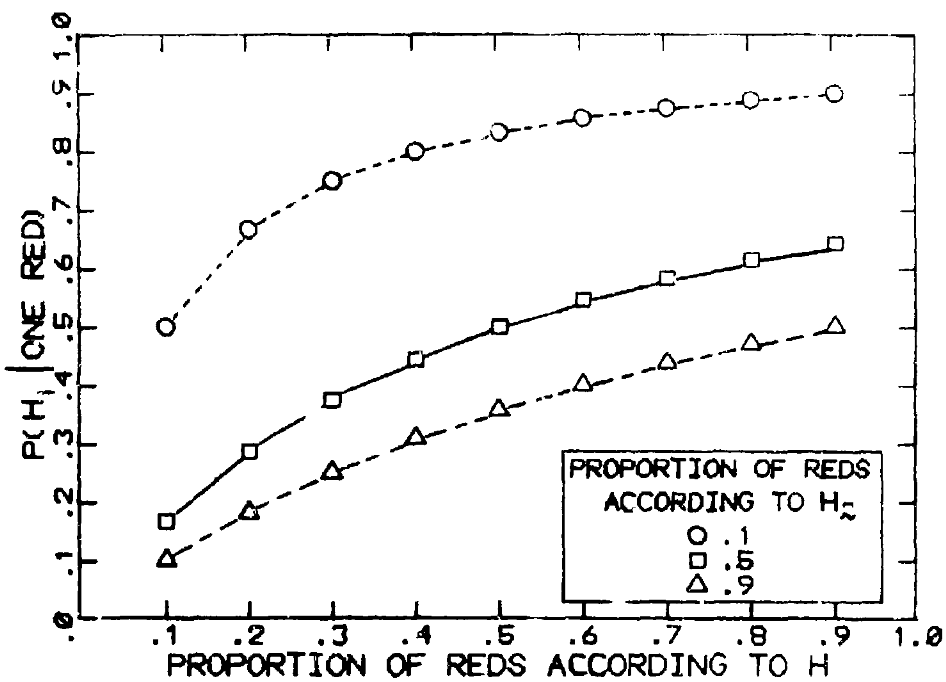
Cells represent expected values of posterior probabilities associated with correct hypothesis after 10 draws from (either) one of the urns.

There are several things to notice about this table. First, it represents the expected value of the probability associated with the true hypothesis, so it represents $p(H_1|D)$ if data are sampled from an urn for which H_1 is true, and $p(H_2|D)$ if the sample is taken from an urn for which H_2 is true. A second thing to notice about the table is the fact that it is symmetric about the minor diagonal. Thus, for example, the probability associated with the true hypothesis is the same for H_1: x% red, H_2: y% red as for H_1: y% red, H_2: x% red. This is a trivial point, and simply indicates that the expected effect of a sequence of observations is strictly a function of the diagnosticity of data and is independent of which hypothesis is which. Third, except when one of the hypotheses is extreme (say, hypothesizes that 10% or less, or 90% or more, of the chips are of one color), the expected impact of data is largely a function of the difference between the hypothesized percentages and relatively independent of their absolute magnitudes.

It was pointed out above that for various combinations of H_1 and H_2, the first one or two observations can have a remarkably large effect. How much effect they will have depends, however, on what those observations are and on the disparity between H_1 and H_2. This point is illustrated by figure 18. The figure shows the probability of H_1, given a single observation that yields a red chip. In all cases, it is assumed that the hypotheses were equally probable before the observation. Note that if the hypotheses are disparate, for example, H_1: 90% red and H_2: 10% red, or H_1: 10% red and H_2: 90% red, a single observation will change the probabilities associated with H_1 and H_2 from .5 and .5 to .9 and .1, or to .1 and .9. On the other hand, if the initial probabilities are very close, say .5 and .6, a single observation will change them very little.

8.6.8 The Symmetrical Two-Hypothesis Case

A two-hypothesis case of special interest is that for which one of two possible observations has the same probability given one hypothesis as does the other observation given the other hypothesis. That is, we are concerned with the situation in which $p(D_\alpha|H_1) = p(D_\beta|H_2)$, or equivalently, in which $p(D_\alpha|H_1) = 1-p(D_\alpha|H_2)$. This is sometimes referred to as the "symmetrical" case, reflecting the fact that one of the two possible observations provides exactly as much support for one of the hypotheses as does the other observation for the other hypothesis. This situation holds in the chips-in-urn context when both hypotheses involve the same proportional split of chips of different colors, but one identifies red chips, and the other black chips, as being the more numerous.



$p_0(H_1) = p_0(H_2) = .5$

Figure 18. The effect of drawing a single Red chip, given various combinations of prior hypotheses concerning the proportion of Reds in the urn

The hypothesis pair H_1: 70% R, 30% B; H_2: 30% R, 70% B satisfies this condition, for example; whereas the pair H_1: 70% R, 30% B; H_2: 20% R, 80% B does not.

The symmetrical case is of special interest because of the fact that the effect of a series of observations on the odds favoring one hypothesis over the other can be calculated in a trivially simple way. If Ω_0 represents the odds prior to the observations of interest, and L represents the likelihood ratio, then

$$\Omega_d = L^d \Omega_0 \tag{25}$$

where d represents the difference between the number of observations of D_α (say, red chips) and of D_β (say, black chips), and Ω_d represents the odds following the observations. Note that the size of the sample--the number of observations--does not enter into this calculation. Suppose, for example, that $\Omega_0 = 1$ and $L = 3$ (as would be the case if $p(R|H_1) = .75$ and $p(R|H_2) = .25$, and the odds and likelihood ratio were expressed H_1 relative to H_2), then, given a sequence of observations yielding four more red chips than black chips, the posterior odds would be

$$\Omega_4 = 3^4 \cdot 1 = 81, \tag{26}$$

and the same result would hold whether the difference of four was obtained from a sample containing 8 reds and 4 blacks or one containing 100 reds and 96 blacks.

The exclusive dependence of Ω_d on d follows directly from the fact that the likelihood ratio for one of the two possible observations is the reciprocal of that for the other observation. Recall from equation (11) that the posterior odds following a single observation is simply the prior odds multiplied by the likelihood ratio associated with the observation

$$\Omega_n = L\Omega_{n-1}. \tag{11}$$

Recall, too, however, that the likelihood ratio is conditional on the observation. Thus, if D_α is observed,

$$L = \frac{p(D_\alpha|H_1)}{p(D_\alpha|H_2)} \tag{27}$$

whereas, if D_β is observed,

$$L = \frac{p(D_\beta|H_1)}{p(D_\beta|H_2)}. \tag{28}$$

Letting L_α represent the likelihood ratio when D_α is observed, and L_β the likelihood ratio when D_β is observed, we may represent the effect of a <u>specific</u> sequence of observations, say $D_\alpha D_\alpha D_\beta D_\alpha D_\beta D_\alpha$, on the odds as

$$\Omega_n = L_\alpha^4 \; L_\beta^2 \; \Omega_{n-6} . \tag{29}$$

In the symmetrical case, however,

$$L_\beta = L_\alpha^{-1} , \tag{30}$$

so the effect of the same specific sequence of observations may be written as

$$\Omega_n = L_\alpha^4 \; L_\alpha^{-2} \; \Omega_{n-6} = L_\alpha^2 \; \Omega_{n-6} , \tag{31}$$

and in general

$$\Omega_n = L_\alpha^d \; \Omega_{n-k} , \tag{32}$$

where d is the number of observations of D_α minus the number of observations of D_β. But, inasmuch as neither n nor k is used in the calculation, we may express Ω as a function of d, and write the expression as in equation (25).

We see then that in the symmetrical case, the odds increase exponentially with the difference between the number of observations of the one type and that of the other type that have been obtained. Figure 19 shows how the rate of growth of this function depends on the disparity between the conditional probabilities, or, equivalently, on the size of the likelihood ratio. Figure 20 shows how the size of the difference that is required to realize a given odds varies with the larger of the conditional probabilities of which the likelihood ratio is comprised. The finding that people typically tend to be conservative Bayesians in their use of data to revise their estimates of the likelihoods of the possible states of the world suggests that many people would find the relationships that are shown in these figures to be counterintuitive. The fact,

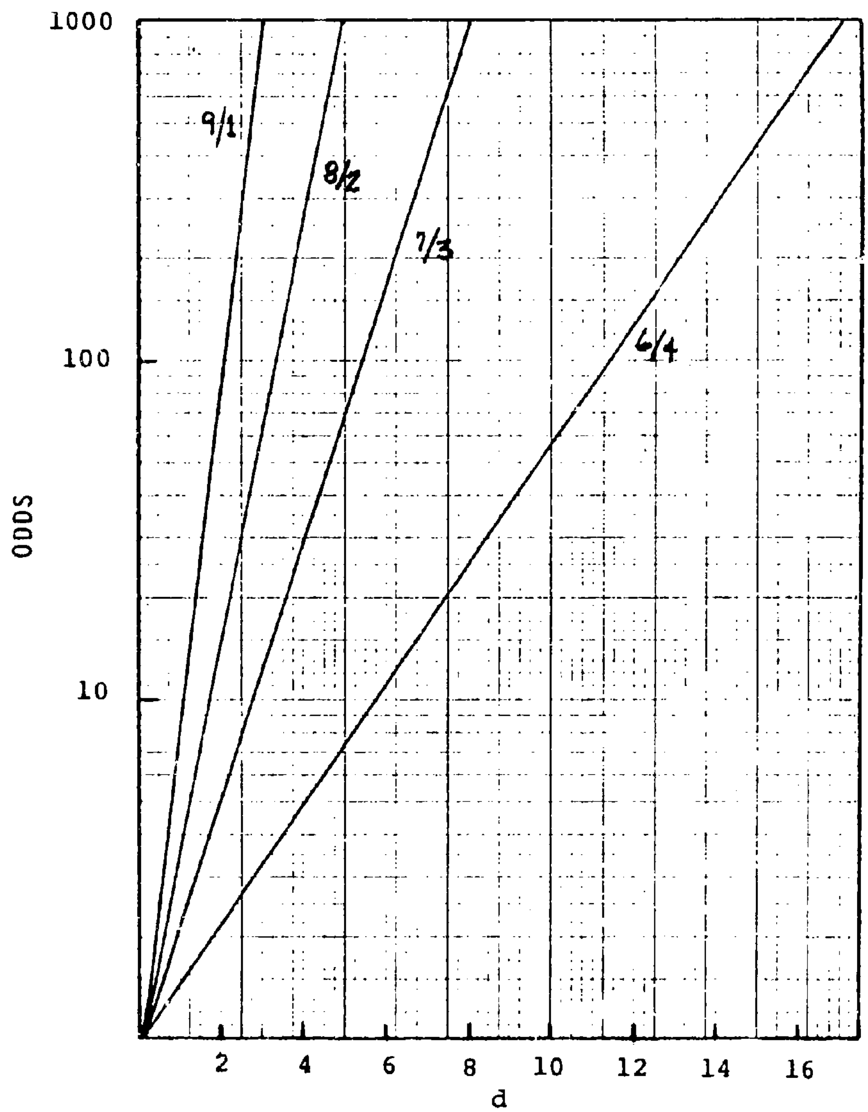


Figure 19. The rate of growth in posterior odds as a function of the difference, d, between the number of observations favoring H_1 and the number favoring H_2, in the symmetrical case (The parameter is likelihood ratio, $L_{1,2}$)

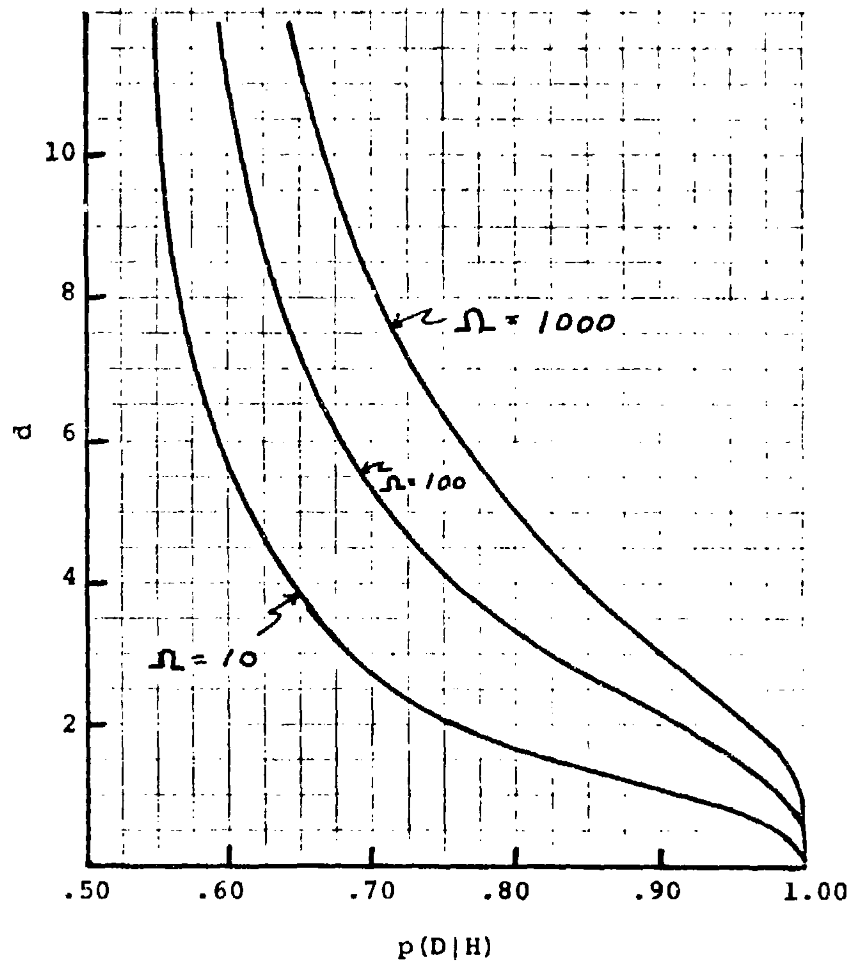


Figure 20. The size of d required to realize a given odds ratio as a function of the larger of the two conditional probabilities in the symmetrical case

for example, that with conditional probabilities of $p(D_\alpha|H_1) = .7$ and $p(D_\alpha|H_2) = .3$, a sample that contains three more observations of D_α than of D_β will favor H_1 over H_2 by a factor of more than 10 may be surprising; as may the fact that with a difference of six, the odds are greater than 100 to 1.

Another aspect of the symmetrical case that some readers may find counterintuitive is the fact that the total effect of a series of observations on the odds depends only on the difference between the number of observations of the two types and is independent of the total number of observations made. Both intuition and statistical training suggest that one's confidence in any inference that is to be drawn from the properties of a sample should increase with the sample size. The apparent paradox is resolved by a recognition of the fact that, except under the hypothesis that each observation is equally likely, the absolute difference (though not the relative difference) between the frequencies of occurrence of the two types of observation is expected to increase with sample size. Specifically, if H: x% R, (1-x)% B is true, the difference between the number of Rs and Bs in a sample of size N should be $(2x-1)N$, on the average.

Consider, for example, the symmetrical hypotheses H_1: 70% R, 30% B and H_2: 30% R, 70% B. If H_1 were true, samples of ten draws would be expected to produce four more reds than blacks on the average; and the odds following a ten-draw sample with four more reds than blacks would be about 30 to 1 in favor of H_1. Samples of 100 draws, given H_1, should produce 40 more reds than blacks, on the average, a difference that would drive the odds to more than 523 trillion to 1. Thus, the odds do tend to increase with sample size because d tends to increase with sample size. A sample of 100 draws that produced four more reds than blacks would be quite unlikely if H_1 were true, and thus would not constitute strong evidence in favor of that hypothesis. It would be even less likely, however, if H_2 were true, so it does constitute some evidence for H_1, but only as much as one would expect to obtain from a much smaller sample. Table 3 shows the odds favoring H_1, given various combinations of $p(D|H_1)$ and $p(D|H_2)$ and several values of d.

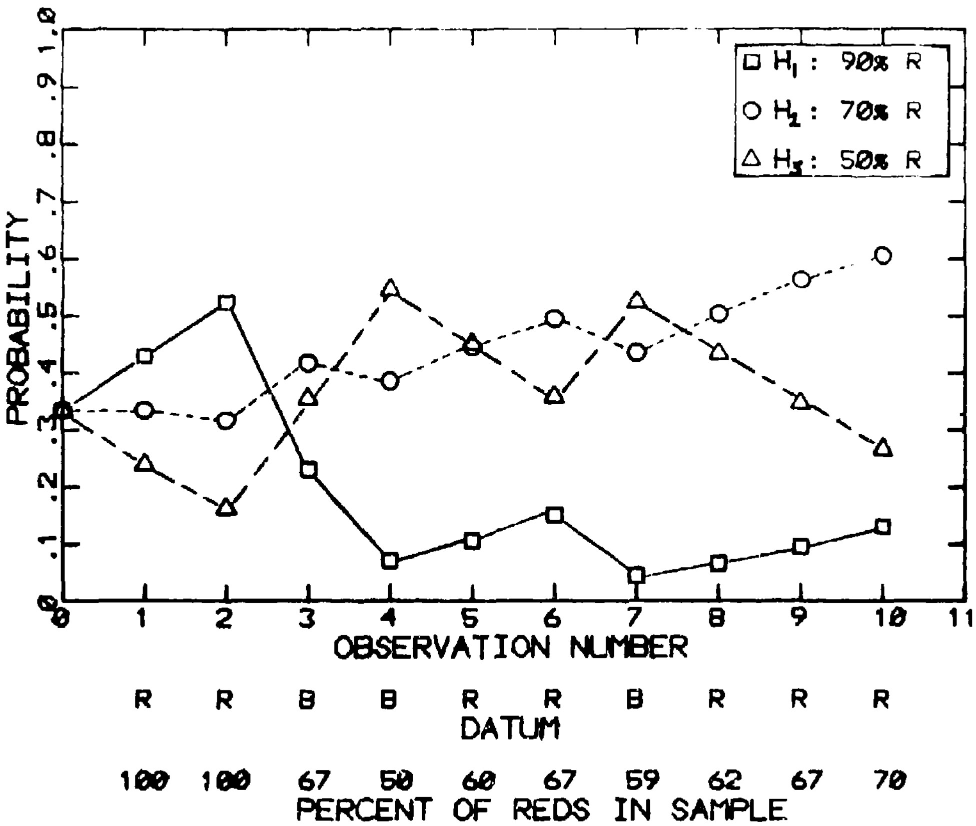
8.6.9 The Several-Hypothesis Case

So far, the examples that we have considered to illustrate the use of Bayes rule have involved only two hypotheses. We turn now to consideration of a few cases in which there are more than two hypotheses. Figure 21 illustrates a case in which H_1, H_2 and H_3 represent, respectively, the hypotheses that the percentage of red chips in the urn is 90, 70 and 50, and shows how the posterior probabilities associated with these hypotheses would change

TABLE 3. ODDS FAVORING H_1 GIVEN THE INDICATED VALUES OF $p(D|H_1)$, $p(D|H_2)$ AND d.

			d						
$p(D\|H_1)$	$p(D\|H_2)$	L	1	2	4	8	16	32
.55	.45	1.22	1.2E0	1.5E0	2.2E0	4.9E0	2.4E1	5.8E2
.60	.40	1.50	1.5E0	2.3E0	5.1E0	2.6E1	6.6E2	4.3E5
.65	.35	1.85	1.9E0	3.4E0	1.2E1	1.4E2	1.9E4	3.5E8
.70	.30	2.33	2.3E0	5.4E0	3.0E1	8.8E2	7.7E5	6.0E11
.75	.25	3.00	3.0E0	9.0E0	8.1E1	6.6E3	4.3E6	1.9E15
.80	.20	4.00	4.0E0	1.6E1	2.6E2	6.6E4	4.3E9	1.8E19
.85	.15	5.67	5.7E0	3.2E1	1.0E3	1.1E6	1.1E12	1.3E24
.90	.10	9.00	9.0E0	8.1E1	6.6E3	4.3E7	1.9E15	3.4E30
.95	.05	19.00	1.9E1	3.6E2	1.3E5	1.7E10	2.9E20	8.3E40

All odds values are rounded to two significant digits and expressed in exponential form. To obtain the approximate value of Ω, multiply the number to the left of the E by ten raised to the power indicated by the number to the right of the E. For example, 4.3E7 = 4.3×10^7).



H_1: 90% Red; H_2: 70% Red;
H_3: 50% Red; $p_0(H_i)$ = .333

Figure 21. Changes in posterior probabilities $p(H_i|D)$ as a result of the indicated observations of Red and Black chips

as a result of the indicated sequence of observations, given $p_0(H_i) = .333$. Figure 22 shows the change in uncertainty concerning which hypothesis is correct as a result of the same sequence of observations.

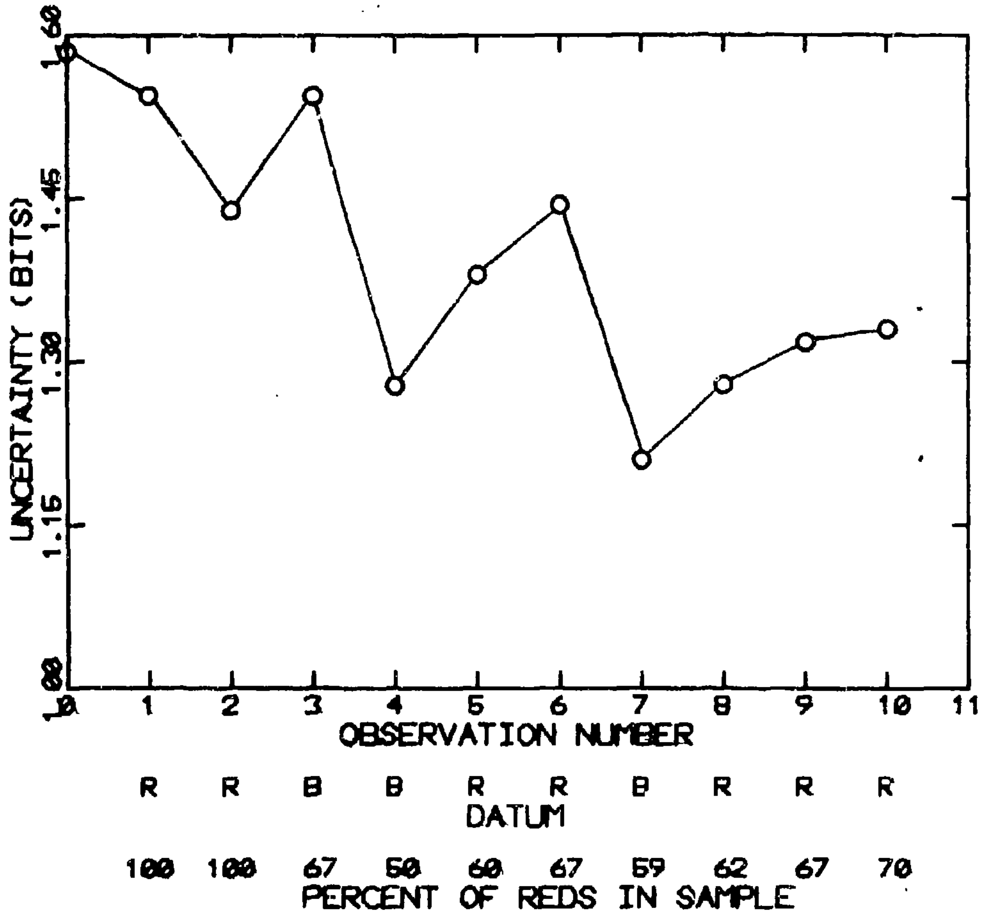
When only two hypotheses are under consideration, there is only one odds ratio (or its reciprocal) that can be expressed. The number of odds ratios that can be expressed grows rapidly, however, as the number of hypotheses is increased beyond two. In general, given N hypotheses, there are $\binom{N}{2}$ or $N(N-1)/2$ odds that can be expressed considering only pairs of hypotheses. Thus in the three-hypothesis case we might consider $\Omega_{1,2}$, $\Omega_{1,3}$ or $\Omega_{2,3}$, each of which is shown in figure 23 for our example.

It may be of interest to consider other than pairwise odds ratios in the several-alternative case, however. Given five hypotheses, for example, one might wish to consider the odds of H_1 relative to the combination of H_3 and H_4, which would be obtained by taking the ratio of $p(H_1|D)$ to the sum of $p(H_3|D)$ and $p(H_4|D)$. It may often be of particular interest to consider the odds of a given hypothesis, H_i relative to all the remaining hypotheses in combination. Such an odds would give the ratio of the probability that H_i is true to the probability that one of the remaining hypotheses is true, i.e., that H_i is false. We might refer to such an odds as the absolute odds of H_i and represent it as follows:

$$\Omega_{i,\bar{i}} = \frac{p(H_i|D)}{\sum\limits_{j,j\neq i} p(H_j|D)} = \frac{p(H_i|D)}{1-p(H_i|D)} \tag{33}$$

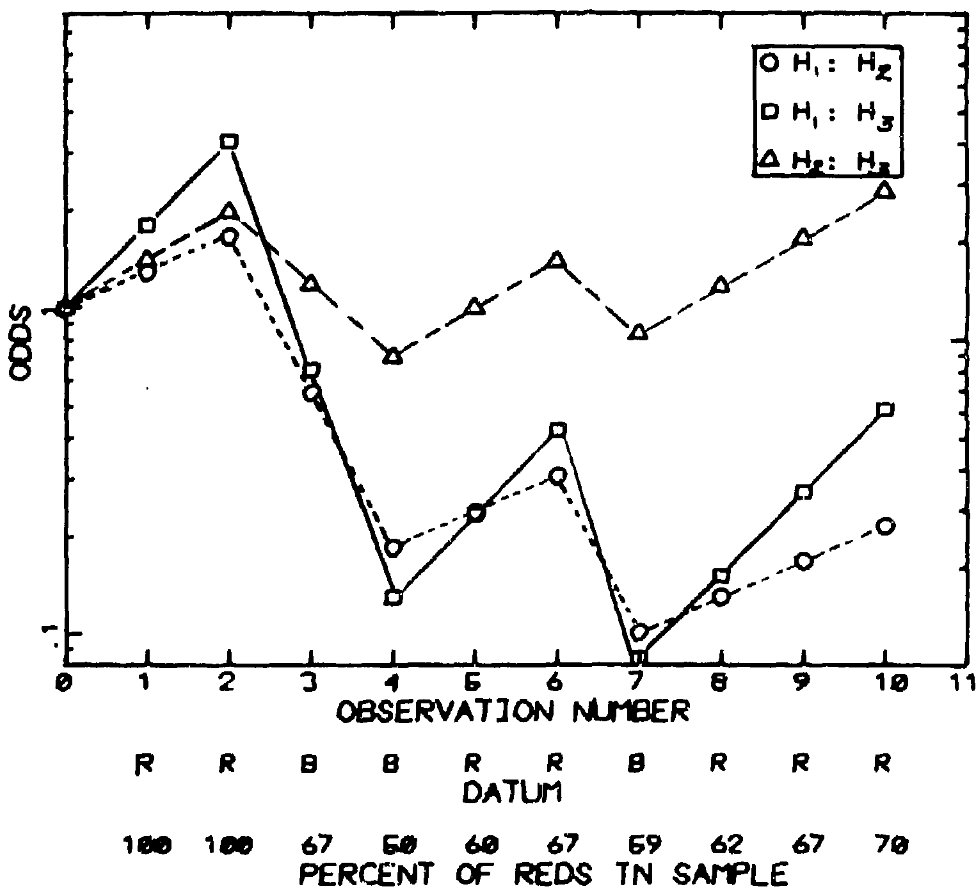
Figure 24 shows how the indicated observations affect the absolute odds of each of the hypotheses of our example.

Expected values of posterior probabilities and of uncertainty may be calculated in the same way when there are several hypotheses as when there are only two. An outcome graph such as those shown in figures 8 and 9 could be used to specify all possible posterior probabilities for a given hypothesis, H_i, and their probabilities of attainment on the assumption that a specified hypothesis, H_j is true. The weighted sum of the nodes above a particular value of N would represent, as before, the expected value, after N observations, of the posterior probability of H_i, given that H_j is really true. Also as before, computation of expected uncertainty involves summing over both i and m, for given N. Inasmuch as it is possible to compute an expectation of $p(H_i|D)$ given that H_j is true for all



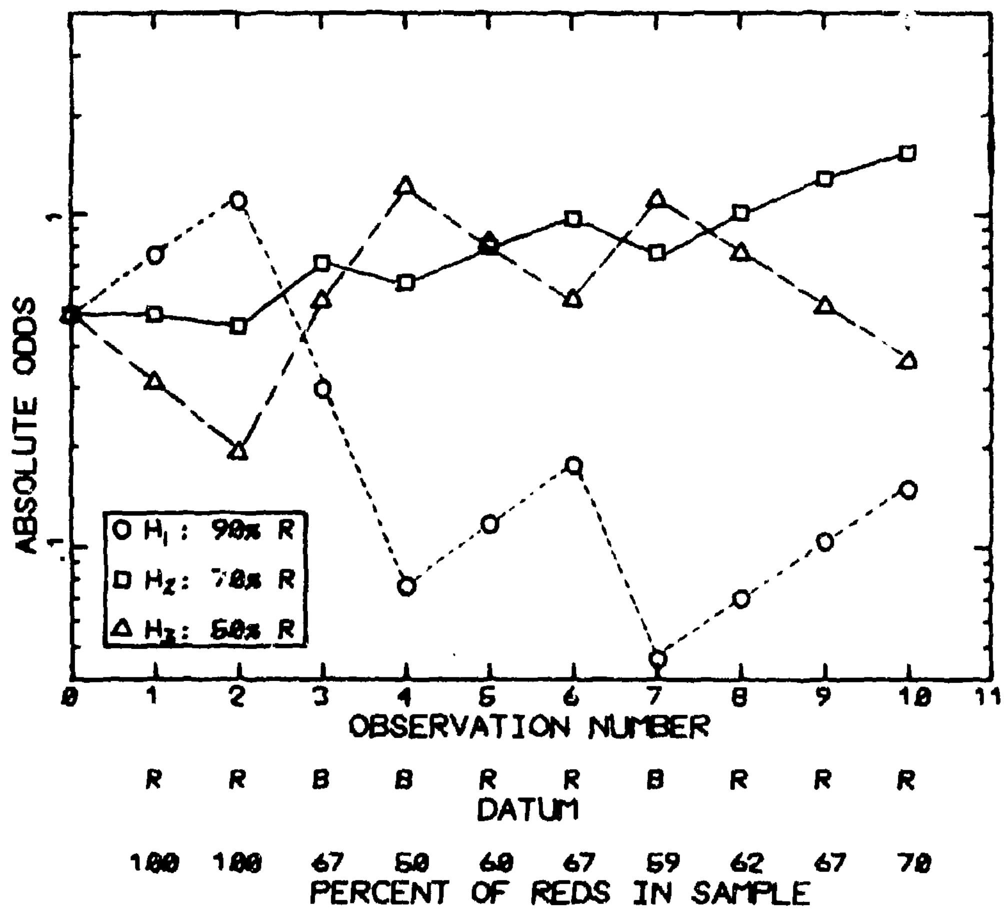
H_1: 90% Red; H_2: 70% Red;
H_3: 50% Red; $p_0(H_i)$ = .333

Figure 22. Changes in uncertainty as a result of indicated observations



H_1: 90% Red; H_2: 70% Red;
H_3: 50% Red; $p_0(H_i)$ = .333

Figure 23. Changes in pairwise odds as a result of the indicated observations



H_1: 90% Red; H_2: 70% Red;
H_3: 50% Red; $p_0(H_i) = .333$

Figure 24. Changes in absolute odds for each hypothesis as a result of the indicated observations

possible combinations of values of i and j, the number of outcome graphs that could be of interest increases with the square of the number of hypotheses under consideration.

Table 4 shows the expected values of $p_N(H_i|D)$ and U_N, for $N = 1,2, \ldots 10$, given that H_j is true for all combinations of i and j in the case of our example (H_1: 90% red, H_2: 70% red, H_3: 50% red, $p_0(H_i) = .333$). As might be expected, given that H_j is true, $E[p(H_i|D)]$ increases most rapidly when $i = j$; which is to say, the expected value of the probability associated with the true hypothesis grows faster than that of the probability associated with either false hypothesis. Counter to intuition, however, this is not a necessary condition. An example will be considered presently in which the expected probability associated with a false hypothesis grows, for a while, at a greater rate, than does the expected probability associated with the true hypothesis, even when both hypotheses are equally probable a priori. With continued sampling, however, the probability of the true hypothesis eventually gets larger than that of any of the false hypotheses. Another point of interest concerning table 4 is the fact that each of three columns of values occurs twice: the second and fourth columns are identical, as are the third and seventh, and the sixth and eighth. This illustrates the following relationship:

$$E[p(H_i|D)|H_j \text{ is true}] = E[p(H_j|D)|H_i \text{ is true}], \qquad (34)$$

that is, the expected posterior probability of H_i, given that H_j is true, is the same as the expected posterior probability of H_j, given that H_i is true. This relationship holds in general, and independently of the number of hypotheses under consideration.

As in the two-alternative case, the rate at which the expected values of the posterior probabilities approach one or zero--and, consequently, the rate at which uncertainty is expected to decrease--depends on the disparity among the hypotheses. The point is illustrated in table 5, which shows all values of $E[p_{10}(H_i|D)|H_j$ is true] for two sets of hypotheses: H_1, H_2 and H_3: 90, 70 and 50% red, and 90, 60 and 30% red. The table also shows the expected uncertainty after ten observations, $E(U_{10})$, concerning which hypothesis is true, as a function of which hypothesis actually is true.

Table 6 shows $E[p_{10}(H_i|D)|H_j$ is true] and $E(U_{10}|H_j$ is true) for two sets of five hypotheses. This table illustrates some of the same points as does table 4. The rate at which the probabilities change from their original values, and the rate at which uncertainty decreases depend on the disparity among the hypotheses. The value of $E[p(H_i|D)|H_j$ is true] is always equal to that of $E[p(H_j|D)|H_i$ is true], which is seen by the fact that each array, if considered as a matrix, is equal to its transpose.

TABLE 4. EXPECTED VALUES OF POSTERIOR PROBABILITIES, AND UNCERTAINTY (IN BITS), GIVEN THAT CHIPS ARE SAMPLED FROM THE URN FOR WHICH THE INDICATED HYPOTHESIS IS TRUE.

	True Hypothesis											
	H_1				H_2				H_3			
	Hypothesis for which Expectation Computed											
Number of Observations	H_1	H_2	H_3	E(U)	H_1	H_2	H_3	E(U)	H_1	H_2	H_3	E(U)
1	.397	.333	.270	1.53	.333	.333	.333	1.49	.270	.333	.397	1.45
2	.453	.327	.220	1.44	.327	.338	.334	1.41	.220	.334	.446	1.35
3	.501	.319	.180	1.35	.319	.347	.334	1.36	.180	.334	.486	1.26
4	.542	.310	.149	1.26	.310	.358	.332	1.32	.149	.332	.519	1.19
5	.577	.300	.123	1.18	.300	.370	.330	1.28	.123	.330	.547	1.12
6	.607	.291	.102	1.10	.291	.383	.326	1.25	.102	.326	.572	1.06
7	.634	.281	.084	1.03	.281	.396	.322	1.22	.084	.322	.593	1.00
8	.658	.272	.070	0.96	.272	.411	.318	1.19	.070	.318	.612	0.95
9	.680	.262	.058	0.90	.262	.425	.312	1.17	.058	.312	.630	0.91
10	.699	.253	.048	0.85	.253	.440	.307	1.14	.048	.307	.645	0.87

H_1: 90% red, H_2: 70% red, H_3: 50% red; $p_0(H_i) = .333$.
(Note: expected uncertainty, E(U), is not the same as the uncertainty calculated from the expected posterior probabilities.

TABLE 5. $E[p_{10}(H_i|D)|H_j$ IS TRUE$]$ FOR ALL COMBINATIONS OF i and j AND THE TWO INDICATED HYPOTHESIS SETS.

H_1: 90% Red, H_2: 70% Red, H_3: 50% Red

		j = 1	j = 2	j = 3
i	1	.699	.253	.048
	2	.253	.440	.307
	3	.048	.307	.645
E(U) (in bits)		0.85	1.14	0.87

H_1: 90% Red, H_2: 60% Red, H_3: 30% Red

		j = 1	j = 2	j = 3
i	1	.824	.171	.005
	2	.171	.603	.226
	3	.005	.226	.769
E(U) (in bits)		0.48	0.86	0.55

In both cases $p_0(H_i) = .333$.

TABLE 6. $E[p_{10}(H_i|D)|H_j$ IS TRUE] FOR TWO FIVE-HYPOTHESIS SETS.

H_1-H_5: 90, 80, 70, 60, 50% Red, respectively

		j=1	2	3	4	5
i	1	.475	.282	.148	.068	.026
	2	.282	.271	.216	.146	.085
	3	.148	.216	.239	.221	.176
	4	.068	.146	.221	.273	.292
	5	.026	.085	.176	.292	.421
E(U) (in bits)		1.64	1.90	1.96	1.86	1.67

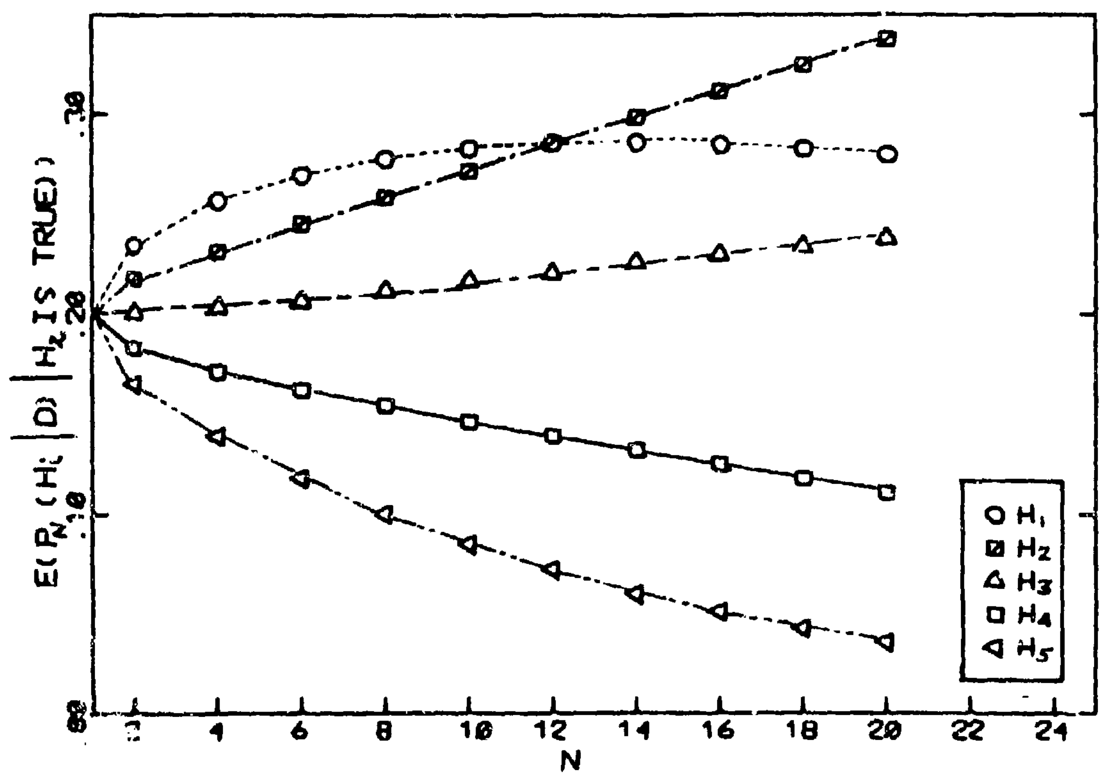
H_1-H_5: 90, 75, 60, 45, 30% Red, respectively

		j=1	2	3	4	5
i	1	.604	.280	.094	.021	.002
	2	.280	.337	.240	.112	.031
	3	.094	.240	.300	.240	.126
	4	.021	.112	.240	.323	.304
	5	.002	.031	.126	.304	.537
E(U) (in bits)		1.19	1.63	1.75	1.63	1.30

In both cases $p_0(H_i) = .2$.

For the hypothesis set represented by the bottom half of table 6, it is true that the expectation is maximum when $i = j$, which is to say that after ten observations the probability of the true hypothesis is always larger than that of any of the false ones. Note, however, that this property does not characterize the values for the hypothesis set represented by the top half of the table. In particular, given this hypothesis set, the expected probability of H_1 is greater than that of H_2 after ten observations, even if chips are drawn from an urn for which H_2 is true. Similarly, $E[p_{10}(H_5|D)]$ is greater than $E[p_{10}(H_4|D)]$ when H_4 is true. With continued sampling the expected probability of the true hypothesis will continue to grow, finally approaching one, whereas that of each of the false hypotheses will at some point begin to decrease and will eventually approach zero. The fact that the expected value of the probability of a false hypothesis is higher at any time than that of the true hypothesis may be quite counterintuitive, however. Figure 25 shows the way in which the expected values of each of the posterior probabilities of the example represented in the top half of table 6 change over twenty observations, given that H_2 is really true. Note that $p(H_2|D)$ is initially smaller than $p(H_1|D)$, but eventually overtakes and surpasses it; with further sampling $p(H_2|D)$ would continue to increase, whereas $p(H_1|D)$ would decrease.

A comparison of tables 5 and 6 illustrates several additional points. The hypothesis sets represented in table 5 are contained within those represented in table 6. Considering only those hypotheses that are represented in both tables, it may be seen that the expected posterior probabilities associated with hypotheses within the smaller set are invariably larger than those associated with the same hypotheses within the larger set. It may also be seen that the expected amount of uncertainty remaining after ten observations, given the truth of a specific hypothesis, is greater when the hypothesis set contains five alternatives than when it contains three. Of course, the a priori uncertainty is also greater in the former case (2.32 bits versus 1.58 bits), so what is of greater significance is the fact that a larger proportion of the original uncertainty is resolved in the three-alternative case.



H_1: 90% Red; H_2: 80% Red;
H_3: 70% Red; H_4: 60% Red;
H_5: 50% Red; $p_0(H_i) = .2$

Figure 25. Expected value of posterior probability of H_i, given that H_2 is true, as a function of number of observations

8.6.10 Man as a Bayesian Hypothesis Evaluator

A considerable amount of experimentation has been done to determine how well Bayes rule predicts behavior when an individual attempts to process probabilistic information in situations like those illustrated. For example, given the task of deciding, on the basis of a sequence of observations, which of several hypotheses about the nature of the source of those observations is true, how closely will the estimates produced by the human decision maker correspond to those produced by the application of Bayes theorem? Obviously, in situations as highly structured as those described, it would be of little interest to do such experiments with an individual who understood Bayes rule and was permitted to do the calculations necessary to use it. Such a test would do nothing but demonstrate one's ability to do arithmetic. Experiments on Bayesian information processing typically are done with people who are not formally aware of Bayes rule, or if they are, they are not provided with the time to perform the necessary calculations. It is an interesting question, in this case, whether an individual's intuitive, or at least informal, notions about evidence will lead him to adjust his probability estimates in a way similar to that that would result from an application of Bayes rule. And if the answer to this question is no, it is of interest to determine whether his performance deviates from that of Bayes rule in consistent ways.

Perhaps the question that has been of greatest interest to, and received most attention from, experimenters is whether hypotheses are more effectively evaluated by having decision makers estimate posterior probabilities, $p(H|D)$, directly upon acquiring incoming data, or to have them estimate conditional probabilities, $p(D|H)$, and then to use these estimates to update the posterior probabilities with the use of Bayes rule. Much of the evidence favors the conclusion that hypotheses are evaluated more efficiently when the latter approach is taken, that is, when humans make estimates of $p(D|H)$ and these estimates are used along with Bayes rule to calculate estimates of $p(H|D)$. Although the directional effects of data on posterior probability estimates produced by humans are similar to those on estimates revised in accordance with Bayes rule, the magnitudes of the effects tend to be smaller in the former case. In particular, the posterior probabilities tend to obtain more extreme values and to reach asymptote faster when they are calculated according to Bayes theorem than when they are estimated directly by humans (Edwards, Lindman, & Phillips, 1965; Howell & Getty, 1968; Kaplan & Newman, 1963; Peterson & DuCharme, 1967; Peterson & Miller, 1965; Peterson, Schneider, & Miller, 1965; Phillips & Edwards, 1966. It appears, therefore, that humans tend to extract less information from data than the data contain; they require more evidence than does a Bayesian process to arrive at a given level of certainty

concerning which of the competing hypotheses is true. That is one of the findings that has led to the characterization of man as a "conservative" Bayesian. In other words, men tend to underestimate high posterior probabilities and overestimate low ones. A similar, but less pronounced, tendency is found when men estimate odds rather than posterior probabilities (Phillips & Edwards, 1966).

Slovic and Lichtenstein (1971) refer to the conservatism of man in his use of probabilistic data as the primary finding of Bayesian research. They review three competing explanations of the result: (1) misperception, or misunderstanding, by the subject of the process by which the data are generated; (2) inability of subjects to aggregate, or put together, the impacts of several data to produce a single response; and (3) an inability, or unwillingness, to assign extreme odds, e.g., odds outside the range of 1:10 to 10:1. Whether any of these explanations is adequate has yet to be determined.

It was the finding of conservatism that prompted Edwards (1963, 1965) and his colleagues (Edwards, Phillips, Hays, & Goodman, 1968) to experiment with probabilistic information-processing systems that use experts to judge the likelihoods of the data reaching the system, given each hypothesis under consideration, and machines to calculate posterior probabilities on the basis of these estimates and the data.

Not all of the evidence that is relevant to the question favors the conclusion that humans are invariably much better at estimating $p(D|H)$ than $p(H|D)$. Southard, Schum, and Briggs (1964b), for example, obtained some results that challenge the generality of the finding that humans tend to underestimate high posterior probabilities, and overestimate low ones. In particular, given a small hypothesis set and a frequentistic environment, the estimates of $p(H|D)$ produced by humans were close to, and sometimes more extreme than, those produced by Bayesian methods. Other studies, several from the same laboratory, have also yielded results that question the validity of the general conclusion that better decisions result when values of $p(H|D)$ are derived by applying Bayes rule to men's estimates of $p(D|H)$ (Schum, Goldstein, & Southard, 1966; Howell, 1967; Kaplan & Newman, 1966; Southard, Schum, & Briggs, 1964a). Often even when evidence of conservatism has been found, the degree to which the human's estimate of $p(H|D)$ has differed from an estimate produced by Bayes rule has been very slight (Peterson & Phillips, 1966; Schum, Southard, & Wombolt, 1969).

These findings do not permit one to conclude that estimates of $p(H|D)$ are <u>never</u> better when derived from estimates of $p(D|H)$ than when <u>produced</u> directly, but they do call into question the opposite notion, namely that of the invariable superiority of the indirect approach. Moreover, they suggest that the direction that research should take is that of determining the conditions under

which each approach is warranted. Schum, Goldstein, and Southard (1966) present some data, for example, that suggest that estimates of $p(H|D)$ that are produced directly are more adversely affected by degradation in the fidelity of the incoming information than are those that are derived from estimates of $p(D|H)$.

Another finding that is relevant to the question of man's capabilities as a Bayesian hypothesis evaluator is that evidence that tends to confirm a favored hypothesis may be given more credence than evidence that tends to disconfirm it (Brody, 1965; Geller & Pitz, 1968; Pitz, Downing, & Reinhold, 1967; Slovic, 1966). This finding raises the more general question of whether a vested interest in a decision outcome impairs one's ability to evaluate data objectively. If it is the case, as Bacon (1955) long ago suggested, that "what a man had rather were true, that he more readily believes," at least one of Savage's basic rules for the application of decision theory is generally violated.

The possibility that an individual's preferences among hypotheses may impair his ability to evaluate them in an unbiased way is closely related to the finding that people tend to be reluctant to change a decision once it has been made (see Section 4.3).

Each of these tendencies--conservatism, partiality, and perseverativeness--has been viewed as a fault, or as evidence that man applies data to the evaluation of hypotheses in an inefficient way. And, in the context of most laboratory experiments in which it has been observed, it undoubtedly is. These tendencies may sometimes be less patently unjustifiable outside the laboratory, however. An insistence on having compelling evidence before changing an established opinion may have a stabilizing effect that is not altogether bad. Many opinions are formed slowly over a period of years, and all the factors that may have contributed to their formation cannot always be recalled at will. The individual who is quick to change an opinion every time he encounters an argument that he cannot immediately refute may find himself constantly shifting from one position to another, always a proponent of the view that he last heard capably expounded.

Hypothesis evaluation has been studied more than most aspects of decision making in the laboratory. This is due in part to the existence of a simple prescriptive model (Bayes rule) for performing this task, given an appropriately structured problem, and in part to the fact that it lends itself to laboratory exploration more readily than some of the other decision-making functions. Much has been learned about man's capabilities and limitations in applying evidence to the resolution of uncertainties about the various aspects of a decision situation. Much remains to be

determined, however. Among several issues that deserve further study are the following: the possibility that information-display formats and response techniques may influence subjective probability estimates that are obtained (Damas, Goodman, & Peterson, 1972; Herman, Ornstein, & Bahrick, 1964); the apparent lack of understanding of how to combine probabilities arising from independent sources of information (Fleming, 1970); the possibility that the weight that one attaches to data may depend on when those data occur during the hypothesis-evaluation process (Chenzoff, Crittendon, Flores, Frances, Mackworth, & Tolcott, 1960; Dale, 1968; Peterson & DuCharme, 1967); and the possibility that one's ability to deal with uncertainty in a conflict situation may depend on whether one is operating with an advantage or a disadvantage with respect to one's opponent (Sidorsky & Simoneau, 1970).

8.6.11 Bayesian Hypothesis Evaluation and Training

One way to interpret some of the results that have been described above--for example, the finding that men often extract less information from data than does a Bayesian aggregator--is to see them as indications that man's intuitive notions concerning the uses of evidence are not entirely consistent with the implications of Bayes rule. Perhaps the thing to do, if this is the case, is to disabuse would-be decision makers of those faulty intuitions.

Such a task might be approached in two ways. On the one hand is the cognitive approach of teaching the decision maker about Bayes rule and its implications. An alternative possibility is to expose the decision maker to a variety of situations, in which his behavior is evaluated and immediate feedback is provided to him concerning the way in which it departs from optimality, if it does. This is the behavior-shaping approach; in essence, it is aimed at modifying one's intuitions without necessarily providing an intellectual understanding of how optimality is defined. These two approaches are not mutually exclusive, of course, and it seems reasonable to assume that a training program would be more likely to be effective if it used both. That is to say, the decision maker should probably be given a good understanding of the notion of inverse probability and how Bayes rule aggregates data; and he should also be provided with considerable practice in attempting to apply the rule in situations that are sufficiently well-structured that his performance can be evaluated and compared to an objective criterion of optimality. The selection of training scenarios should put special emphasis on those situations for which man's intuitions have been shown to be most misleading, e.g., especially small or especially large levels of a priori uncertainty and situations in which the direction of evidence changes after a tentative decision has been reached.

The results of some studies have indicated that such training can be at least partially effective. Fleming (1970), for example, explored the question of the effectiveness of feedback concerning the outcome of a selected action in improving the decision maker's performance on subsequent decision tasks. The context of the study was a simulated tactical decision-making situation. Subjects were required to combine probabilistic data from three independent sources in order to arrive at an estimate of the relative likelihood of attack on each of three ships. Initially, subjects demonstrated an ignorance of the proper combining rule (multiplication) and were conservative in estimating the overall probabilities of attack. The investigator concluded that these data-aggregation and probability-estimation tasks should be automated. He also showed, however, that, although the subjects were unable to generate the correct probabilities on the basis of feedback, they did revise their estimates over the course of trials in such a way as to correct for conservatism (apparently by adding a constant).

Other investigators have also shown that experience in estimating posterior probabilities can produce behavior which, if not optimal, is more nearly so than before the training began (Edwards, 1967; Hoffman & Peterson, 1972; Southard, Schum, & Bridges, 1964b). Such studies establish that certain aspects of hypothesis evaluation, in particular posterior probability estimation, can be improved somewhat as a result of practice. What they do not indicate, however, is how much can be expected of training or how the training should be done in order to obtain optimal results.

Another issue that relates to training involves the question of how well people can make the $p(D|H)$ judgments that they are required to make in some Bayesian systems. It seems to be generally assumed that people have less trouble making these judgments than they do making judgments of $p(H|D)$. In at least one study, however, this was not the case. Bowen, Feehrer, Nickerson, Spooner, and Triggs (1971) encountered a fairly strong resistance on the part of experienced military intelligence officers to the idea of making judgments of the sort: "If it is assumed that 'Attack' is the enemy commander's course of action, what is the probability that one will observe the traditional indication 'Massing of Tanks'?" These investigators pointed out that the "generally negative reaction to the possibility of estimating probabilities of the type that would be required in a Bayesian system must be tempered by the fact that the participants were not familiar with the concept of Bayesian inference and had not been trained to make the required judgments" (p. 103). There is, therefore, the question of the degree to which training in Bayesian analysis would be effective in overcoming the relatively strong preferences that some decision makers seem to have for estimating posterior probabilities themselves.

Several other questions concerning man's capabilities as they apply to hypothesis evaluation have been noted above. These questions have arisen because of the results of experimental studies. They are questions which, for the most part, have not yet been adequately answered. The questions, in most cases, suggest some limitation or deficiency in man's hypothesis-evaluation skills. To the extent that these limitations or deficiencies are demonstrated by further research to be genuine, they represent challenges to designers of training programs. If it is the case, for example, that probability estimations are sensitive to the format in which information is displayed or the mode in which the response is given, as some studies have indicated, the question is whether such effects can be eliminated by training. If they cannot be, then the need to be restrictive with respect to display formats and response modes is so much the greater. Or, to take another example, if the way one applies data to the evaluation of a hypothesis is different for a favored hypothesis than for an unfavored one, as other studies have suggested, this constitutes another challenge to training. Can one be trained to apply data to all possible hypotheses in an unbiased way without regard for his preferences? Similar questions concerning the potential effectiveness of training can be raised concerning each of the other limitations and deficiencies that have been noted. More research will be required in order to answer these questions.

8.7 The Measurement of Subjective Probability

Throughout this report we have made frequent reference to subjective probabilities, and it has been tacitly assumed that such things can be accurately measured. In fact, how to assure accuracy in measurements of this quantity has been a question of some interest. The problem is a problem because of the fact that the probabilities that one obtains may depend on the way in which they are obtained; or as Toda (1963) puts it, subjective probability is essentially defined by the measuring technique that is used. Toda further suggests several criteria that such a measuring technique should satisfy: "First, the logical nature of the task presented to the subject should be thoroughly understood by the experimenter, and, hopefully, by an intelligent subject. Second, the task should involve well-defined payoffs to the subject. Third, the task should be so structured that it is to the disadvantage of a subject to respond in a manner inconsistent with his expectations. Fourth, since our interest in measuring subjective probability is related to its use in decision theory, the measurement technique should not be inconsistent with decision theory" (p. 1).

The third of these criteria is perhaps the most subtle, and has received the greatest amount of attention. Stated in other terms, the requirement is that it be in the subject's best interest to state his probability estimates honestly. That this can be a problem may be illustrated by a simple example of a situation in

which the requirement is not met. Consider the case of a student taking a multiple-choice examination. Suppose he has been instructed that in answering each question he is to assign a number to each of the alternatives associated with that question in such a way as to reflect his estimation of the probability that that alternative is the correct one. When he is very certain of which alternative is correct, all of the numbers except one will be zero; when he is less than 100% certain, however, he would assign nonzero numbers to more than one alternative. Suppose further that the score that he is to receive for any given question is some linear function of the ratio of the number placed on the correct alternative to the sum of the numbers used on all the alternatives associated with that question. Given this scoring rule, the student should not distribute numbers in accordance with his true estimation of the probabilities; instead, he should put zeros on all the alternatives except the one that he considers most likely, even if he is not very certain that that alternative is indeed the correct one.

This is easily seen by considering a two-alternative case. Suppose that the student really thinks that the chances are 7 in 10 in favor of A being the correct alternative. If he is honest, then he will assign 7/10, of whatever points he is going to use, on alternative A and 3/10 on B. Given our scoring rule, and assuming that our hypothetical student assigns numbers to the two alternatives in the ratio of 7 to 3, then the two values that his score may assume are 7/10 and 3/10. Moreover, from the student's point of view, the probability of getting a score of 7/10 is 7/10 (i.e., the probability that A is correct), and the probability of getting a score of 3/10 is 3/10. Thus, the subjectively expected value of his score is $(7/10)^2 + (3/10)^2 = .58$. But suppose that our student were a gambler, and decided to put all his chances on the alternative that he considered most likely to be correct. Now the two values that his score can assume are 1 and 0, and the expected value of his score (assuming that he really believes that A's chances are 7 in 10, rather than 10 in 10, as his answer would indicate) is $7/10 \times 1 + 3/10 \times 0 = .70$. Thus, whereas the student was instructed to assign numbers to alternatives in accordance with his judgment of the likelihood of their being correct, the scoring rule is such that he can expect to obtain a higher score by ignoring the instructions than by following them.

A scoring rule that is to satisfy Toda's "honesty is the best policy" requirement must have what has been referred to as a "matching property." In formal terms, the matching property may be stated as follows: Suppose that a subject reports n non-negative values,

$$r_1, r_2, \ldots r_n, \sum_{i=1}^{n} r_i > 0,$$ presumably to reflect the

subjective probabilities that he associates with alternative possibilities, $x_1, x_2, \ldots x_n$. Assume a discrete subjective

probability distribution, $p_1, p_2, \ldots p_n$, that represents the subject's true probability estimates regarding $x_1, x_2, \ldots x_n$. Letting $\underline{P}$, $\underline{R}$ and $\underline{X}$ represent, respectively, the vectors $(p_1, p_2, \ldots p_n)$, $(r_1, r_2, \ldots r_n)$ and $(x_1, x_2, \ldots x_n)$, and $W(\underline{R}, \underline{X})$ the payoff to the subject, given the response vector $\underline{R}$ and the probability vector $\underline{X}$, the matching property is realized by any payoff function for which the following statement is true: The response vector, $\underline{R}$, maximizes the subjectively expected payoff $E[W(\underline{R}, \underline{X})]$, if and only if $\underline{R} = k\underline{P}$, k being a scalar constant. That is to say, a payoff scheme, or a scoring rule, has the matching property if and only if the subject maximizes his subjectively expected payoff when the weights that he assigns to the possibilities differ from his true subjective probabilities at most by the same multiplicative factor. Note that when the relationship $\underline{R} = k\underline{P}$ does hold, the calculation of odds will be the same whether based on $\underline{R}$ or on $\underline{P}$.

Subjective-probability measurement procedures and response scoring techniques that make use of functions that have this matching property have been referred to as "admissible probability measures" (Shuford, Albert, & Massengill, 1966), and "proper scoring" rules (Winkler & Murphy, 1968). Several functions with the matching property have been defined and investigated, among them the "logarithmic loss" (Good, 1952; Toda, 1963), the "quadratic loss" (Brier, 1950; deFinetti, 1962; Toda, 1963; van Naerssen, 1962, and the "spherical gain" (Toda, 1963; Roby, 1964, 1965).

8.7.1 The Logarithmic Loss Function

The logarithmic loss function is unique among these functions in its exclusive dependence on the value of the component of $\underline{R}$ that is assigned to the correct alternative. It is not affected by how numbers are distributed over the other components of $\underline{R}$. The function is given by

$$W_L(\underline{R}|x_i) = k \log r_i - \sum_{j=1}^{n} r_j \qquad (35)$$

where k is a positive constant, and $(\underline{R}|x_i)$ is read "response vector $\underline{R}$, given that x_i is the correct alternative. The subjectively expected payoff, given this function, is

$$E(W_L) = k\Sigma p_i \log r_i - \sum_{j=1}^{n} r_j \qquad (36)$$

which is maximized when $r_i = p_i$,

$$\text{Max } E(W_L) - k\Sigma p_i \log p_i - 1 \quad \text{(Toda, 1963).} \qquad (37)$$

Because the maximum subjectively expected value is negative--hence its designation as a "loss" function--a constant is often added to the function in order to make the payoff positive. Also, because the function becomes $-\infty$ at $r_i = 0$, a truncated version of it is usually employed in practice.

8.7.2 The Quadratic Loss Function

The quadratic loss function is given by

$$W_Q(R|x_i) = \sum_{k \neq i} r_k^2 + (1 - r_i)^2 \tag{38}$$

when the number of alternatives is greater than two, and by

$$W_Q(R|x_i) = -r_k^2, \; k \neq i \tag{39}$$

for the two-alternative case. This function is negative in the two-alternative case (although not necessarily when the number of alternatives is greater than two), so, as in the case of the logarithmic loss, a constant is often added to the function to assure a positive payoff.

8.7.3 The Spherical-Gain Function

The spherical-gain function, which has been elaborated by Roby (1965) will be considered in somewhat more detail, because it has some useful properties that the other rules do not have, and a particularly elegant geometrical representation as well. The payoff function is given by

$$W_S(\underline{R}|x_i) = r_i \left(\sum_{j=1}^{n} r_j^2 \right)^{-\frac{1}{2}} . \tag{40}$$

For a proof that W_S is maximized only when $\underline{R} = k\underline{P}$ see Shuford, Albert, and Massengill (1966). A reference to the example that was used earlier should be sufficient to make the assertion plausible. Consider again the two-alternative examination item for which a student thinks the chances are 7 in 10 in favor of alternative A. Recall that if his score is a linear function of the proportion of points he assigned to the correct alternative, then his best strategy is to put zero on every alternative except the one he considers most likely to be correct; in which case, his expected score would be .70. To see that this is not true in the case of the spherical gain scoring rule, note that if the student puts all his stakes, say n points, on alternative A, his expected score will be:

$$7/10 \times (n/\sqrt{n^2+0^2}) + 3/10 \times (0/\sqrt{n^2+0^2}) = .70.$$

If, however, he weights the alternatives in accordance with his judgment of what the chances really are, his expected score will be:

$$7/10 \times (7/\sqrt{7^2+3^2}) + 3/10 \times (3/\sqrt{7^2+3^2}) = .76.$$

It should be noted that the procedure permits the student to assign weights to the various alternatives in any way he sees fit. There might appear to be some advantage in forcing the numbers assigned to the alternatives for a given item to add to one, inasmuch as they could then be interpreted directly as probability estimates. The student could be instructed to make his assignments so that they would indeed add to one; however, this is an unnecessary demand since the score is unaffected by a change of scale. Moreover, if we wish to treat the assignments as probability estimates, as we shall in what follows, we can easily normalize them by simply dividing each assigned number by the sum of the numbers associated with that question. When this is done, and each of the original numbers is replaced with the resulting quotient, then each of the resulting numbers will be referred to as a probability estimate, and the collection of numbers associated with a given item as a probability vector.

A nice feature of the spherical gain scoring rule is that it provides an easy and intuitively meaningful way of distinguishing between one's confidence in the truth of a particular hypothesis (or correctness of a test item) and one's general degree of "resolution" with respect to the overall decision space (or to the whole test item). Roby defined, as a "resolution index,"

$$RI = \left[\sum_{j=1}^{n} \left(\frac{r_j}{\sum_{i=1}^{n} r_i} \right)^2 \right]^{\frac{1}{2}} \qquad (41)$$

where RI represents an individual's confidence in his answer. Equation 41 is simply the denominator of equation 40 after the latter has been normalized.

As in the case of W_s, the maximum value of RI is 1. It should be clear that RI = 1 only if $r_j = 1$ for one value of j and 0 for all others. That is to say, in keeping with our intuitive notions about how an index of confidence should behave, it assumes its maximum value when one has put all his chances on a single alternative. (Note that whether that alternative is correct or

incorrect is irrelevant to this measure--as it should be.) Unlike W_s, RI cannot assume the value 0. Its minimum value depends on the number of alternative hypotheses under consideration, or--in the case of the examination example--the number of candidate answers associated with a question. It is obtained when $r_j = \frac{1}{\Sigma r_j}$

for all j; that is, the index gets its lowest value when the same number is assigned to every alternative. Again, this is consistent with our intuitive ideas about confidence. The fact that the minimum value of the index depends on the number of alternatives is also in keeping with our intuitions about how a measure of confidence should behave: one should have less confidence in a guess among three equally likely alternatives than in a guess between two of them.

8.7.4 Implementation of Admissible Probability Measures

One of the practical difficulties in applying scoring rules with the matching property is that of providing subjects with intuitively meaningful information concerning the implications of their probability assignments vis-a-vis the scores that could result from them. It is clear that simply providing individuals with formal expressions of the rules will not suffice, at least for those who are not mathematically trained. One approach to this problem is that of illustrating the implications of any given rule with concrete examples that make clear the advantages of being honest. Another, and perhaps preferred, approach is that of providing the individual with an explicit representation of the payoff that he would receive, given the truth of any specific hypothesis and the way in which he had distributed probabilities over the alternatives.

Organist and Shuford designed a paper and pencil procedure for providing this information in the case of the logarithmic loss function (Baker, 1964; Organist, 1964; Organist & Shuford, 1964). Shuford (1967) and Baker (1968) have also described a computer-based technique for providing similar information in a dynamic way. In this case the alternatives open to the decision maker are shown on a computer-driven display. Associated with each alternative is a line, the length of which represents the user's relative confidence that that alternative is the correct one. The user adjusts the lengths of the lines by means of a light pen. When the length of one line is changed by the user, the lengths of all the others are adjusted by the computer so as to constrain the sum of the lengths to add to one at all times. Also displayed with each line is a number which indicates to the user what his payoff would be if the alternative associated with that line were the correct one. The logarithmic loss function determined the relationship between the number representing

potential payoffs and the lengths of the lines in the applications of the system that are reported. However, the relationship could as well have been determined by any other scoring rule of interest.

8.7.5 The Efficacy of Admissible Probability Measures

It would seem clear from the mathematics of the situation that scoring rules that have the matching property should be used in preference to those that do not. It has not been clearly demonstrated empirically, however, that subjects tend to behave dishonestly if such rules are not used, or that their responses are free of biases if they are (Aczel & Pfanzagl, 1966; Jensen & Peterson, 1973; Samet, 1971; Schum, Goldstein, Howell, & Southard, 1967). Moreover, it is also apparent that many of the probability estimation situations of interest to investigators of decision making are situations in which the only scoring rules that are operative are those that are imposed by nature. The situations in which subjective probabilities are of greatest practical significance tend to be those in which the payoffs are beyond the experimenter's control.

8.7.6 Subjective Probability Measurement and Training

One question of interest that relates to training research is whether individuals who have had experience in making probability judgments in controlled situations with scoring rules that have the matching property are more effective at judging probabilities in real-world situations than those who have had experience at estimating probabilities but have not been exposed to matching-property rules. As has already been noted, some investigators have advocated the use of experts to estimate conditional probabilities to be used in Bayesian aggregation systems (Bond & Rigney, 1966; Edwards, 1965b). Often, however, it is not possible to determine how accurately such estimates are made. If one had an objective indicant of the probabilities of interest that was independent of the experts' judgments, it would not be necessary to get the judgments. It would be of interest, however, to determine whether the behavior of experts on such tasks would be sensitive to the type of experience they had had in estimating probabilities in controlled situations, and in particular to their exposure to admissible or inadmissible probability measurement techniques. Savage (1971) has suggested the early introduction of admissible scoring rules to children, along with careful training in the assessment of opinion strength, could have the salutory effect of dispelling some of the myths concerning the relationships between certainty, belief and action--e.g., the idea that one should speak and act as though certain, even when one is not, nd the notion that weakly held opinions are worthless--that are fostered by conventional educational testing methods.

Scoring rules with the matching property answer to one aspect of the problem of measuring subjective probabilities, namely that of structuring the situation so that honesty in reporting is the best policy. There are other aspects of the problem, however, that are not so readily solved. Expressions of certitude have been shown to vary considerably as a function of the way in which they are reported (Samet, 1971) and of the context in which they are obtained (Nickerson & McGoldrick, 1963). Typically, when subjects are asked to rate their confidence in their own performance on a perceptual or cognitive task, a positive correlation between these variables is found--confidence is highest when performance is best--but the strength of the relationship is not always great, and the significance of a given confidence rating depends on the situation and the person making it (Andrews & Ringel, 1964; Nickerson & McGoldrick, 1965). A fundamental question that is raised by these results is whether such factors affect certitude itself, or only its expression. A further question is whether such variability--whatever its basis--can be eliminated, or at least significantly reduced, as a result of appropriate training.

8.8 The Use of Unreliable Data

In the foregoing discussions of the use of Bayes rule, it has been tacitly assumed that the data used in estimating conditional or posterior probabilities had been accurately observed and reported. In the chips-in-urn illustrations, for example, it was assumed that one could examine a chip and determine its color easily, or that someone else determined the color and reported it accurately. Thus, the decision maker could operate with complete confidence in the data at his disposal. In the real world of decision making, things often are not this way. Frequently, the observation or the reporting of events is faulty, and the decision maker is obliged to take this fact into account when making use of the data that he has obtained.

We naturally assume that data from a trustworthy source will be more useful to a decision maker than will data from a source that has not inspired confidence in the past. The use of an explicit reliability rating procedure for intelligence reports by NATO army forces (see Section 5.2) is based on such an assumption. Few attempts have been made, however, either to validate this assumption or to determine in a quantitative way exactly how confidence in a data source does affect the way in which the data from that source are applied to a decision problem.

8.8.1 Prescriptive Approaches

One class of prescriptive models for taking into account the reliability of data has come to be known as "cascaded" or "multistage" inference, suggesting a process of hypothesis evaluation that involves more than one step. Schum and DuCharme (1971) point out that research on cascaded inference has been focused on two

situations: one, in which the observer or reporter of an event expresses his degree of certainty concerning whether or not the event actually occurred (see, for example, Dodson, 1961; Gettys & Willke, 1959; Steiger & Gettys, 1972), and a second, in which the report of an event is made without qualification by a source that is known to be less than perfectly reliable (see, for example, Schum & DuCharme, 1971; Schum, DuCharme, & DePitts, 1973; Snapper & Fryback, 1971).

In both cases, attention has been confined primarily to relatively simple situations, e.g., those in which (1) the decision maker's task is to determine which of two hypotheses, H_1 or H_2, is true, and observations have only two possible outcomes, D_1 and D_2, and (2) the reliability of a report is independent of the hypotheses that are being considered, that is to say, event D_1 and D_2 are neither more nor less likely to be confused under H_1 than under H_2.

Dodson (1961) considered the situation in which an observer is not certain which of two mutually exclusive events, D_1 and D_2, has occurred, but may be able to make a probability or certitude judgment on the question. He suggested that in order to calculate the posterior probability of a hypothesis in this case, one should calculate its value, given each of the possible events, and then take a weighted sum of these values, the weights being the probabilities that the observer attaches to the event possibilities. Given only two possible data, the calculation may be represented as follows:

$$\xi(H_i|D) = \psi(D_1)p(H_i|D_1) + \psi(D_2)p(H_i|D_2) \tag{42}$$

where $\xi(H_i|D)$ is the posterior probability of H_i, taking the observer's uncertainty into account, and $\psi(D_j)$ is the probability that the observer attaches to the possibility that he has observed event D_j. More generally, given n possible events and the assumption that the observer can attach a probability to each of them, the formula might be written as

$$\xi(H_i|D) = \sum_{j=1}^{n} \psi(D_j)p(H_i|D_j). \tag{43}$$

Substituting the Bayesian formula for $p(H_i|D_j)$ we have

$$\xi(H_i|D) = \sum_j \psi(D_j) \frac{p(D_j|H_i)p(H_i)}{\sum_i p(D_j|H_i)p(H_i)}. \tag{44}$$

Using Dodson's work as a point of departure, Gettys and Willke (1969) and Schum and DuCharme (1971) gave the process of dealing with unreliable data a more explicit two-stage form. The following

discussion roughly follows Schum and DuCharme. These writers focused on the case in which a decision maker obtains information about data via a source that sometimes incorrectly reports what has actually occurred. (It is irrelevant to this discussion whether the source's errors are assumed to be errors of observation or errors of report.) For each of the decision problems that were analyzed there were two hypotheses, H_1 and H_2, two possible data events, D_1 and D_2, and two possible reports by the source, d_1 and d_2. What one wants to determine is $p(H_i|d_j)$.*

According to Bayes rule

$$p(H_i|d_j) = \frac{p(d_j|H_i)p(H_i)}{p(d_j)}. \tag{45}$$

The problem then is to determine $p(d_j|H_i)$. If the probability of a datum conditional on a hypothesis, $p(D_k|H_i)$, and the probability of a report, conditional jointly on a hypothesis and a datum $(p(d_j|H_i \cap D_k))$, are known, then the probability of a report, conditional on a hypothesis $p(d_j|H_i)$ can be easily calculated. The relationship is given by

$$p(d_j|H_i) = \sum_k p(D_k|H_i)p(d_j|H_i \cap D_k), \tag{46}$$

a graphical representation of which is shown in figure 26. When, by assumption, the reliability of a report is independent of the hypothesis that is being considered,

$$p(d_j|H_i \cap D_k) = p(d_j|D_k) \tag{47}$$

so, in effect,

$$p(d_j|H_i) = \sum_k p(D_k|H_i)p(d_j|D_k). \tag{48}$$

Schum and DuCharme refer to

$$\Lambda = \frac{p(d_j|H_i)}{p(d_j|H_k)} \tag{49}$$

*Our notation differs slightly from that used by Schum and DuCharme: we use D_1 and D_2 to represent the two possible data events, whereas they used D and $\bar{D}$, and we use d_1 and d_2 to represent reported data whereas they used D^* and $\bar{D}^*$.

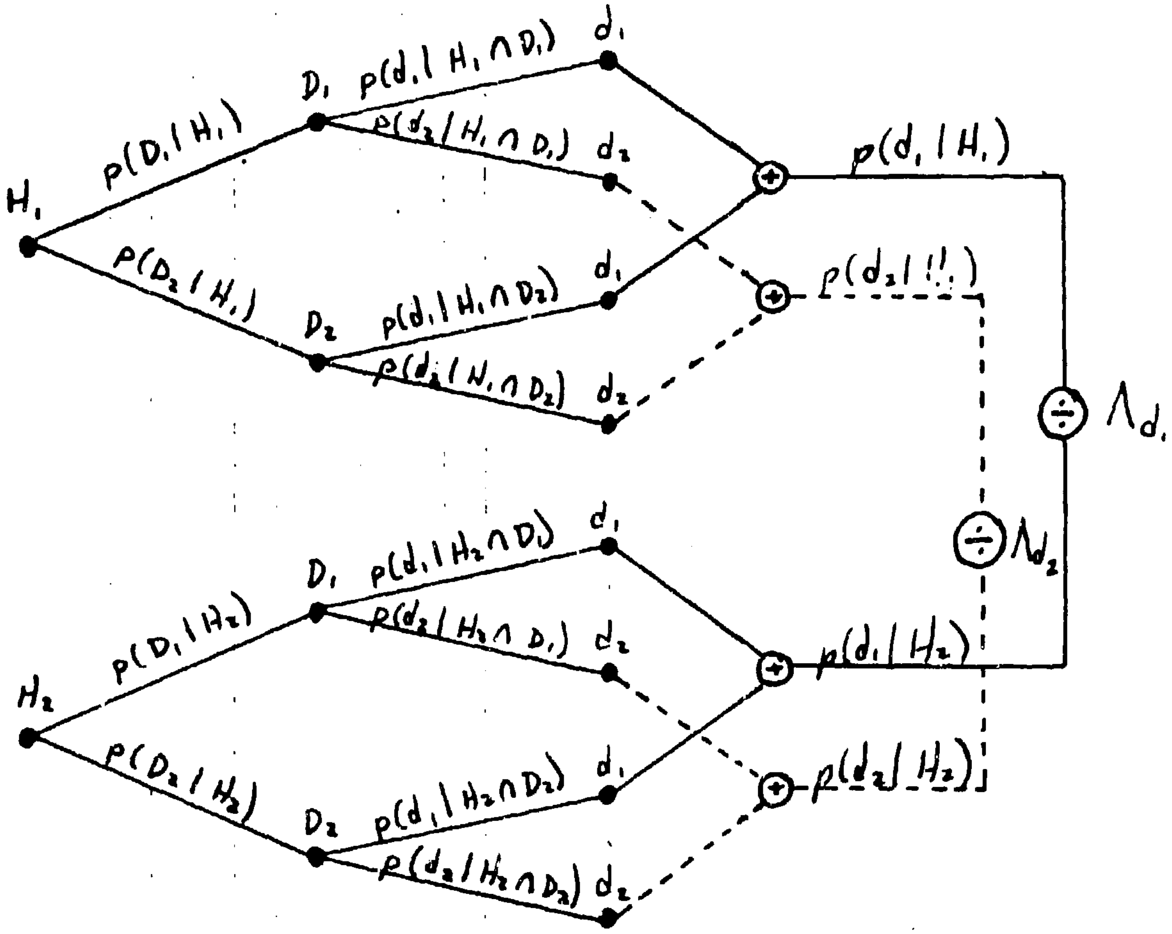


Figure 26. Graphical representation of derivation of $p(d_j|H_i)$ and adjusted likelihood ratios for less than completely reliable data

as the "adjusted likelihood ratio," the likelihood ratio that takes into account the degree of reliability of the source. Of course, Λ reduces to the standard likelihood ratio when the source is assumed to give completely reliable reports, inasmuch as, in this case

$$p(d_j|D_k) = \begin{cases} 1 \text{ for } j = k \\ 0 \text{ for } j \neq k. \end{cases}$$

The way to deal with the problem of unreliable data then, according to Schum and DuCharme is with a two-step process: (1) adjust the diagnosticity of the data by determining $p(d_j|D_k)$ or Λ, and (2) apply the adjusted data to revise the distribution of probabilities over the hypotheses via Bayes rule.

What one must be able to measure or estimate in order to use this procedure are $p(D|H)$, the standard conditional probabilities of Bayes theorem, and $p(d|D)$, the indices of source reliability. Schum and DuCharme define source reliability in terms of

$$r = p(d_i|D_i)$$

the probability that the source will report a data event accurately. They distinguish four different decision "cases" in terms of certain symmetries and asymmetries involving $p(d|H)$ and $p(d|D)$, and they develop the implications of their prescription for dealing with unreliability for each case. The cases that they distinguish are:

Case I: Symmetric $p(D|H)$: Symmetric $p(d|D)$

$p(D_1|H_1) = p(D_2|H_2)$; $p(d_1|D_1) = p(d_2|D_2)$.

Case II: Asymmetric $p(D|H)$; Symmetric $p(d|D)$

$p(D_1|H_1) \neq (D_2|H_2)$; $p(d_1|D_1) = p(d_2|D_2)$.

Case III: Symmetric $p(D|H)$: Asymmetric $p(d|D)$

$p(D_1|H_1) = p(D_2|H_2)$; $p(d_1|D_1) \neq p(d_2|D_2)$.

Case IV: Asymmetric $p(D|H)$; Asymmetric $p(d|D)$

$p(D_1|H_1) \neq p(D_2|H_2)$; $p(d_1|D_1) \neq p(d_2|D_2)$.

In order to avoid the use of conditional probability notation, Schum and DuCharme introduced the following notational equivalencies:

For symmetric $p(D|H)$:

$$p \equiv p(D_i|H_i)$$
$$1-p \equiv p(D_j|H_i),\ j \neq i.$$

For asymmetric $p(D|H)$:

$$p_1 \equiv p(D_1|H_1)$$
$$p_2 \equiv p(D_1|H_2)$$
$$1-p_1 \equiv p(D_2|H_1)$$
$$1-p_2 \equiv p(D_2|H_2).$$

For symmetric $p(d|D)$:

$$r \equiv p(d_i|D_i)$$
$$1-r \equiv p(d_j|D_i),\ j \neq i.$$

For asymmetric $p(d|D)$:

$$r_1 \equiv p(d_1|D_1)$$
$$r_2 \equiv p(d_2|D_2)$$
$$1-r_1 \equiv p(d_2|D_1)$$
$$1-r_2 \equiv p(d_1|D_2).$$

Letting the subscripts on Λ represent symmetry or asymmetry with respect to $p(D|H)$ and $p(d|D)$, respectively, and making the above substitutions into equation (46), as appropriate, we obtain Schum and DuCharme's expressions for the prescribed use of data of imperfect, but known, reliability for each of the four cases they considered. All adjusted likelihood ratios represent

$$\frac{p(d_1|H_1)}{p(d_1|H_2)}.$$

Case I: $$\Lambda_{s,s} = \frac{pr+(1-p)(1-r)}{(1-p)r+p(1-r)} \quad (50)$$

Case II: $$\Lambda_{a,s} = \frac{p_1 r+(1-p_1)(1-r)}{p_2 r+(1-p_2)(1-r)} \quad (51)$$

or, equivalently,

$$\Lambda_{a,s} = \frac{p_1+k}{p_2+k} \quad (52)$$

where

$$k = \frac{1-r}{2r-1},\ r \neq .5 \ . \quad (53)$$

Case III: $$\Lambda_{s,a} = \frac{pr_1+(1-p)(1-r_2)}{(1-p)r_1+p(1-r_2)} \quad (54)$$

or if $p \neq 1$ and $r_2 \neq 1$,

$$\Lambda_{s,a} = \frac{c\left[\frac{p}{1-p}\right]+1}{c+\left[\frac{p}{1-p}\right]} \quad (55)$$

where

$$c = \frac{r_1}{1-r_2} \ . \quad (56)$$

Case IV: $$\Lambda_{a,a} = \frac{p_1 r_1+(1-p_1)(1-r_2)}{p_2 r_1+(1-p_2)(1-r_2)} \quad (57)$$

or if $r_1 \neq (1-r_2)$

$$\Lambda_{a,a} = \frac{p_1+b}{p_2+b} \quad (58)$$

where

$$b = \left[\frac{1-r_2}{r_1-(1-r_2)}\right] \ . \quad (59)$$

It follows from the definitions of unadjusted and adjusted likelihood ratio that the latter is always closer to unity than the former and that the difference between them increases as reliability, r, is decreased from 1.0 to 0.5* (except when the data are completely uninformative to begin with and the unadjusted ratio is 1). This is consistent with the intuitively compelling requirement that the less reliable the data, the less diagnostic impact it should have.** Figure 27 shows how the difference between unadjusted and adjusted likelihood ratio grows as reliability is decreased, and how the adjusted ratio goes to 1 for $r = .5$, for the case in which both $p(D|H)$ and r are symmetric, i.e., Schum and DuCharme's Case I.

Figure 27 also illustrates the fact that the greater the diagnostic impact of data (when reported by a completely reliable source), the greater is the effect of a decrease in reliability of a report. This also is an intuitively reasonable relationship: the less informative data are to begin with, the less there is to lose if they are reported unreliably. What is less intuitively apparent is the fact that even a very small decrease in reliability may have an extremely large effect on likelihood ratio if the unadjusted ratio is very high. Schum and DuCharme (1971) point out, for example, that in Case I, if a datum with an unadjusted likelihood ratio of 100,000 is reported by a source with a reliability of .99, the adjusted ratio is reduced by about four orders of magnitude to slightly less than .99.

The results of Schum and DuCharme's analysis bear on issues relating to the design of information and decision-making systems and on the role of humans therein. For example, they show that under Case I conditions, there is a reasonably straightforward tradeoff

*Decreasing r below 0.5 has the effect of making the adjusted likelihood ratio depart again from unity, although it still remains closer to unity than does the unadjusted ratio. In other words, decreasing the reliability quotient below 0.5 increases the diagnosticity of the data, but in support of the alternative hypothesis. This is consistent with the idea that a source that is consistently wrong may be very informative; one need only interpret its report as evidence of the opposite of what it says. In this discussion, we will confine our attention to the case in which $1.0 \geq r \geq 0.5$.

**Schum and DuCharme (1971) point out, however, that when the reliability of report is not independent of which hypothesis is being considered, it is possible for Λ to differ more from 1 than does L; that is to say, it is possible for a decrease in reliability, in that case, to increase the diagnosticity of data.

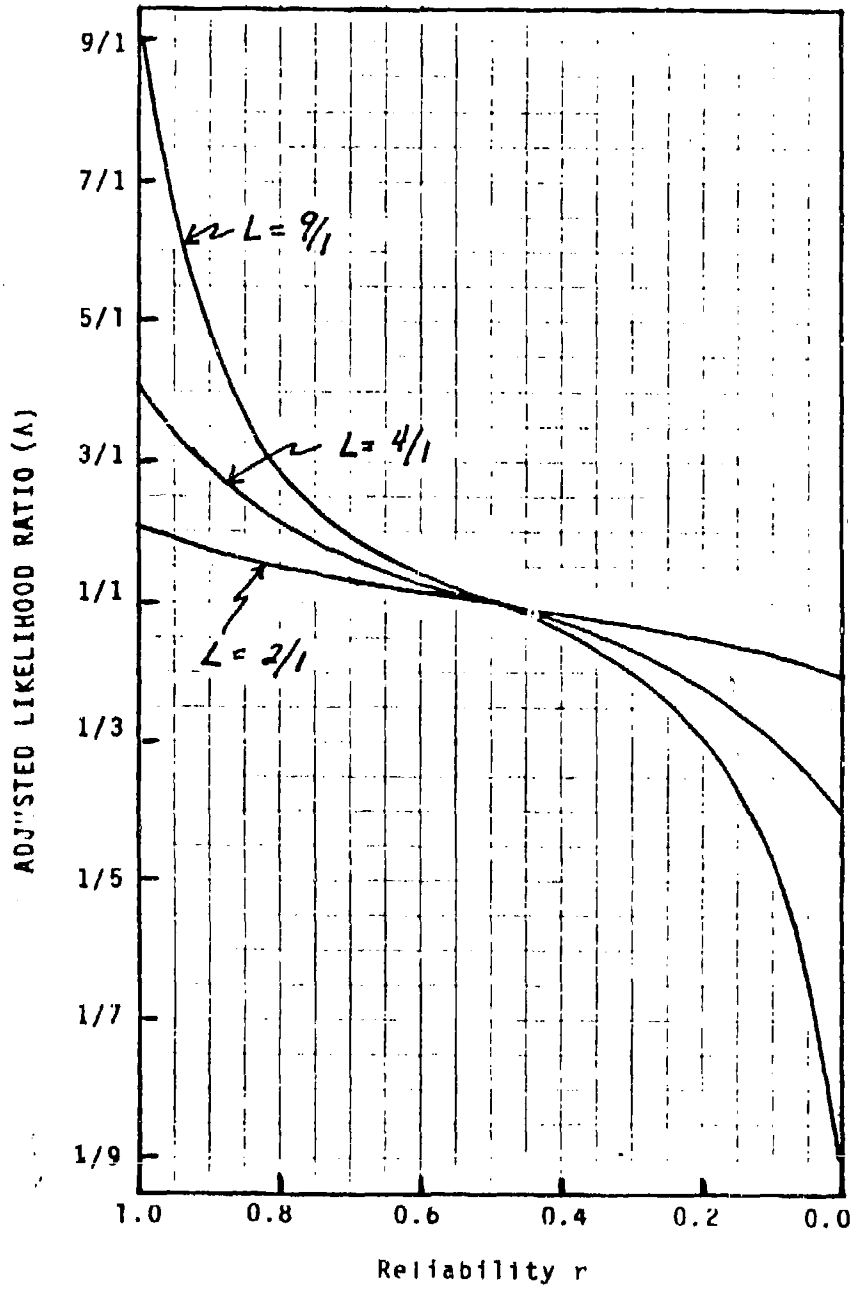


Figure 27. Adjusted likelihood ratio (Λ) as a function of data reliability (r) for several values of unadjusted likelihood ratio (L), for Schum and DuCharme's Case I

between $p(D|H)$ and $p(d|D)$. And the tradeoff is such that if one wants to increase the diagnostic impact of information flowing through a system, and the costs of increasing the conditionals $p(D|H)$ and $p(d|D)$ are equal, one should increase the smaller of the two.

Also, the analyses show that in Cases II and IV Λ is dependent upon specific values of p_1 and p_2 rather than on their ratio. Thus, despite the fact that earlier results have suggested that people find it easier to make judgments of likelihood ratios than of conditional probabilities, there may be situations in which estimates of the latter should be required.

8.8.2 Some Empirical Results

The models developed by Schum and DuCharme are prescriptive, providing for optimal adjustment of the likelihood ratio under conditions in which data are reported with less than total, but known, reliability. We now turn to a consideration of several studies aimed at comparing actual performance against that prescribed by these models. In the next section we then present a brief account of some descriptive models suggested by these results. All experiments and models that will be considered in these sections address situations where input to the decision process is an event or set of events reported by a single unreliable source.

Snapper and Fryback (1971) present the results of a study in which the experimenter reported to the subject with (symmetric) reliabilities of 1.0, 0.9, and 0.7 the outcomes of events conceptually similar to the draws of chips from an urn. The probabilities of events conditional on hypotheses, $p(D_1|H_1)$, $p(D_2|H_1)$ and $p(D_1|H_2)$, $p(D_2|H_2)$, were, respectively, as follows: (a) 0.33, 0.67 and 0.67, 0.33; (b) 0.80, 0.20 and 0.60, 0.40; (c) 0.90, 0.10 and 0.45, 0.55; (d) 0.25, 0.75 and 0.75, 0.25. For conditions in which the experimenter's reliability was equal to unity, only (a) and (b) were used. Subjects were required to indicate which of the hypotheses they considered more likely as a result of the experimenter's report, and how much more likely than the alternative hypothesis they considered it to be. Under conditions of unit reliability, subjects' estimates corresponded very closely to the actual likelihood ratio, but when reliability was less than unity they represented slight underestimates of the impact of the least diagnostic reports and overestimates of the impact of the remaining reports. The extent of this overestimation, moreover, increased with the magnitude of Λ.

Johnson (1974; see also Johnson, Cavanagh, Spooner, & Samet, 1973) has utilized a similar task and response structure to study the effects of four different variables on cascaded inference: (1) sample size, the number of draws which underlay a cumulative outcome report (e.g., "three reds, two blacks");

(2) data generator diagnosticity, the relative composition of red and black chips in the urn; (3) sample diagnosticity, the diagnostic value defined by the difference between total numbers of red and of black chips underlying a report; and (4) source reliability. Posterior-odds estimates that were obtained in this case were found to be sensitive to different values of sample size, data generator diagnosticity and source reliability, tending to decrease as the values of these variables decreased. When the report was known to be perfectly reliable, estimates of posterior odds were generally more conservative than those computed from Bayes theorem; however, they became progressively less conservative and approached optimal values at intermediate levels of reliability (.8-.7), and then became slightly excessive at lower levels (.7-.6).

The diagnosticity and reliability of reported events were manipulated by Youseff and Peterson (1973) in such a way that the value of Λ in a situation requiring multistage inference was equal to the value of the standard likelihood ratio in a single-stage situation (that is, one with report reliability equal to unity). Subjects' estimates tended to be conservative for high values, both of Λ and of L, as compared with the Bayesian model, and tended to be excessive for low values. The odds estimated in conditions requiring multistage inference were consistently greater than those estimated in single-stage conditions and, as a result, were excessive compared to the optimal odds over a wider range than were single-stage odds.

Schum, DuCharme, and DePitts (1971) conducted a study in which the accuracy of subjects' own observations of the number of Xs contained in tachistoscopically presented 4 x 4 matrices of Xs and Os constituted the reliability levels. Subjects were required to estimate the relative likelihood of two possible hypotheses relating to the data generator after each of five stimulus presentations. Under conditions in which sufficient time was available for totally accurate observation of the stimuli, estimates became increasingly conservative compared to the optimal model as the diagnosticity of each observed event and the inferential consistency over a set of five events increased. Under conditions in which insufficient time was available for accurate observations, the subjects' estimates were generally close to optimal or slightly excessive when diagnosticity and consistency were high, and became more conservative as either of these parameters assumed lower values. In a second phase of this same study, subjects estimated directly the diagnosticity of data based on brief observations of each slide. Compared with the optimal model, such estimates become increasingly excessive as L increased.

The results of these studies establish that the behavior of decision makers is indeed influenced by the degree of reliability of their data sources. They also demonstrate, however, that

performance tends not to be consistent with that prescribed by the formally appropriate rule for adjusting data diagnosticity. Further, performance with unreliable data often differs in one important respect from that that has been observed in classical Bayesian inference situations in which events are observed, or reported, with perfect accuracy. Whereas in the latter case the decision maker's estimates, though revised in the appropriate direction, tend to be conservative as compared with Bayes theorem, his estimates based on less-than-completely-reliable data frequently appear to be excessive as compared with Schum and DuCharme's prescription for optimality. Because the value of Λ as defined by the Schum and DuCharme model, in effect, makes an adjustment in the direction of increasing conservatism (produces a value closer to unity), the two effects--conservatism vis-a-vis L and excessiveness vis-a-vis Λ--can offset each other, if conditions are just right.

8.8.3 Some Attempts to Develop Descriptive Models of Cascaded Inference

As we have noted, the model developed by Schum and DuCharme (1971) for dealing with unreliable data prescribes two steps, or stages: in the first stage, the nominal diagnosticity of a datum is discounted to reflect the degree of reliability of the source, and in the second, the adjusted datum is applied to the hypotheses under evaluation in accordance with Bayes rule. If hypotheses are being evaluated in terms of odds, the process can be represented as follows:

$$\text{Stage 1:} \quad \text{compute } \Lambda = \frac{p(d|H_i)}{p(d|H_j)}$$

$$\text{Stage 2:} \quad \text{compute } \Omega_1 = \Lambda\Omega_0$$

where Λ represents the adjusted likelihood ratio, and Ω_1 and Ω_0 represent the posterior and prior odds, respectively.

The experimental results that were reviewed briefly above make it clear that people typically do not behave in accordance with this prescription. Several investigators have attempted to develop models that do describe behavior.

The results obtained by Snapper and Fryback (1971), using symmetric reliabilities, suggest that in dealing with unreliable data, decision makers estimate the likelihood ratio as though the data were completely reliable, adjust the resulting ratio by multiplying it by the reliability quotient, and then apply the adjusted ratio to the calculation of posterior odds. The process may be represented as follows:

Stage 1: compute $\hat{\lambda} = rL$

Stage 2: compute $\hat{\Omega}_1 = \hat{\lambda}\Omega_0$.

Snapper and Fryback note that the optimal rule for the first stage of the process is neither apparent nor intuitive, whereas the rule that seemed to describe the behavior of their subjects has some intuitive appeal and is easily applied. Its use leads, however, to subjective estimates of likelihood ratio that are excessive in comparison with those prescribed by Λ. That is to say, $\hat{\lambda}$ leads to overestimation of the diagnostic impact of a given (unreliable) datum.

The extent to which $\hat{\lambda}$ overestimates Λ--for Schum and Ducharme's Case I--is illustrated in figures 28 and 29. Figure 28 shows both $\hat{\lambda}$ and Λ as functions of r for several values of L; figure 29 shows the ratio $\hat{\lambda}/\Lambda$ for the same conditions. The figures show only cases for which $L \geq 1$ and $r \geq .5$. For $L < 1$, one obtains the same relationships by simply expressing the likelihood ratio H_2 re H_1 rather than H_1 re H_2. The case of $r < .5$ is of little interest for the reason explained in the first footnote on page 134. As may be seen from these figures, the degree to which $\hat{\lambda}$ overestimates Λ depends both on L and r: for given L it tends to vary inversely with r (given $r \geq .5$) and for given r it increases sharply with L.

Gettys, Kelly, and Peterson (1973) have suggested a model that is slightly different from that of Snapper and Fryback. It assumes that the decision maker estimates posterior <u>odds</u> on the assumption that the most likely event is true, and <u>then</u> adjusts the odds to reflect the reliability of the data source. This model may be represented as follows:

Stage 1: compute $\Omega_1 = L\Omega_0$

Stage 2: compute $\hat{\Omega}_1 = r\Omega_1$.

It is apparent that although the process by which the posterior odds are estimated differs in the two cases, the results are precisely the same. Edwards and Phillips (1966) have presented evidence, however, suggesting that the way in which people estimate posterior odds may be better described by

$$\Omega_1 = L^c\Omega_0, \tag{60}$$

where c varies with L, than by the prescribed $\Omega_1 = L\Omega_0$. Funaro (1974) points out that the models of Snapper and Fryback, and of

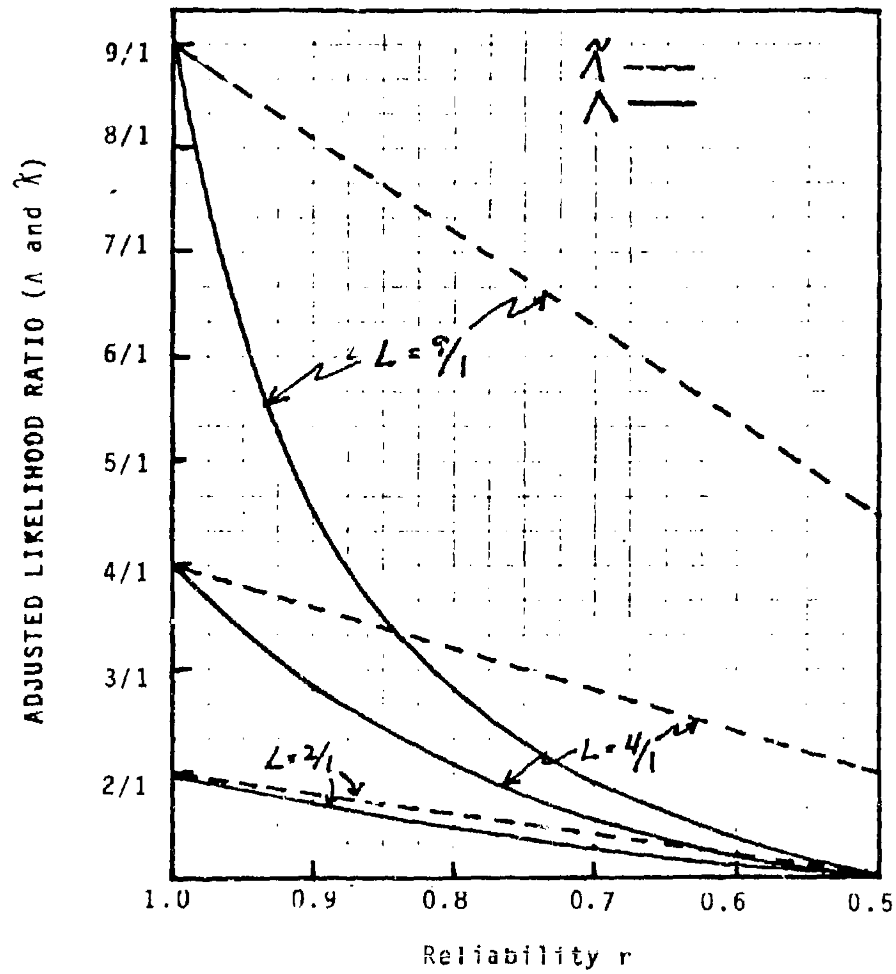


Figure 28. $\tilde{\lambda}$ and Λ as functions of r for several values of L

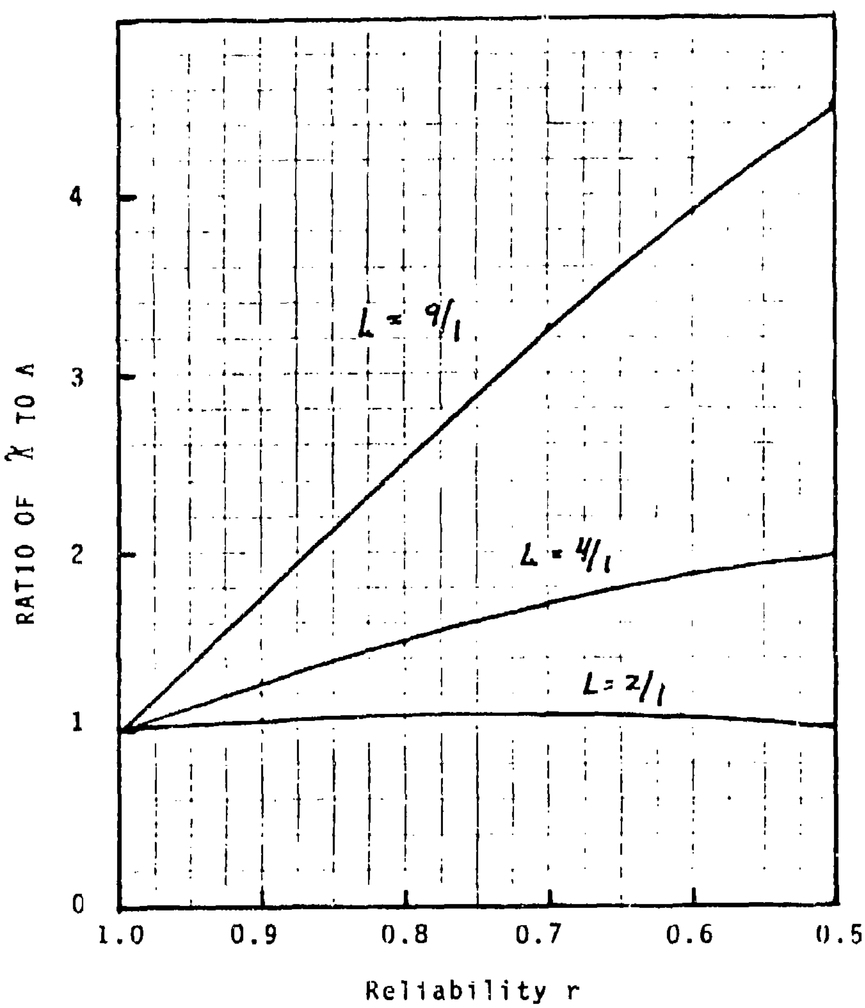


Figure 29. The ratio of λ to Λ as a function of r for several values of L

Gettys, Kelly, and Peterson make different predictions if the odds are calculated according to Phillips and Edwards' expression. The former leads to

$$\tilde{\Omega}_1 = (rL)^c \Omega_0 \qquad (61)$$

and the latter to

$$\tilde{\Omega}_1 = rL^c \Omega_0 . \qquad (62)$$

Funaro (1974) has recently attempted to evaluate the predictive power of Snapper and Fryback's model and of that of Gettys, Kelly, and Peterson, using both L and L^c as unadjusted likelihood ratios in each case. A symmetric p(D|H)--symmetric r task (Schum and DuCharme, Case I) was used. Subjects were required to revise odds' estimates under both single-stage (perfect-source reliability assumed) and cascaded-inference conditions. Values of c were estimated separately for individual subjects from data obtained in the single-stage conditions.

The results were not consistent with any of the models described above. They were predicted best by another model that Funaro proposed. This model, which Funaro called the empirical model, assumes that subjects accurately estimate Λ, and then apply this estimate to the revision of odds with the same degree of effectiveness, or ineffectiveness, with which they apply L in single-stage tasks. The conclusion appears to be inconsistent with the results of Youssef and Peterson (1973) who found that odds's estimates made under cascaded conditions were consistently excessive relative to those made in single-stage tasks, given Λ = L. Funaro notes, however, that subjects in his experiment could have acquired a direct appreciation for Λ from the proportion of successes and failures in a series of reports obtained from the source during the course of the experiment. (In a symmetrical p(D|H) chips-in-urn situation, one can unambiguously define a "success" as the drawing--or in this case reporting--of a chip of the predominant color.) To the extent that subjects were able to develop a direct awareness of Λ, the effect would have been to eliminate the need for a two-stage process and to transform the task into the simpler problem of revising odds on the basis of totally reliable data. The suggestion is an eminently plausible one and the possibility that this is in fact the way unreliable data are often accommodated in real-world situations deserves further study.

8.9 Some Comments on Bayesian Hypothesis Evaluation

Inasmuch as the Bayesian approach to hypothesis evaluation has received so much attention by decision theorists and investigators of decision making, it seems important to consider some of

the limitations of this approach. To point out limitations is not, of course, to deny that the approach has merit. Among its advantages are the fact that it places minimal demands on memory because data can be discarded after being used to update the distribution of probabilities over hypotheses, the fact that it provides a means of aggregating qualitatively different data in a meaningful way and the fact that the procedure for applying data to the evaluation of hypotheses automatically weights data in terms of their relevance to the hypotheses being evaluated. It is precisely because the approach does work well in some contexts that there is a danger of uncritically concluding that it is appropriate in all cases. The following observations are based largely on a discussion by Bowen, Nickerson, Spooner, and Triggs (1970).

First, Bayes rule itself applies to only one of the several aspects of decision making; namely, hypothesis evaluation or, more precisely, the resolution of uncertainty concerning the state of the world. Whatever its efficacy for that particular task, it is not the grand solution to the problem of decision making.

Second, application of Bayes rule requires that the decision problem be structured in a very precise way. In particular, it requires that one's uncertainty about the state of the world be represented as an exhaustive set of mutually exclusive possibilities. It does not, however, provide any help in identifying these possibilities.

Third, the requirement for an exhaustive set of mutually exclusive hypotheses about the state of the world precludes the possibility of expanding one's hypothesis space as one goes along. It clearly often is the case, in real-life situations, that new hypotheses are suggested by incoming data. That is to say, observations may have the effect not only of modifying the credibility of existing hypotheses, but of suggesting new hypotheses as well.

Fourth, the fact that use of Bayes rule presupposes a set of mutually exclusive hypotheses has another implication. By definition, one and only one of these hypotheses can be true; all the others must be false. The probabilities that are associated with these hypotheses do not, of course, represent their truth values, but, rather, the decision maker's opinion concerning their truth or falsity. It was pointed out in the preceding paragraph that no provision is made for the possibility that the hypothesis set does not contain the true hypothesis. It is also the case that provision is not made for the possibility that more than one of the hypotheses are true, or that one or more is partially true.

Fifth, application of Bayes rule is a recursive process: each time that a new observation is to be used to update a posterior probability estimate, the posterior probability from the preceding

update is used as the prior probability for the current update. Originally, however--before the first observation is made--the prior probabilities must be estimated, and Bayes rule does not help in this regard. Investigators are not entirely agreed on how these prior probabilities should be assigned--or on what they mean. It is often pointed out that how prior probabilities are assigned may make little difference (provided values very close to 0 or 1 are not used), because the effect of the initial values will be largely nulled after several observations have been made. The problem can be a significant one, however, when hypotheses must be evaluated on the basis of relatively few data. In such cases, the initial prior probabilities can have a very strong effect on the final posteriors, and thus the way in which they are assigned is of considerable concern.

Sixth, the basic assumption that justifies a Bayesian approach to hypothesis evaluation is the assumption that man is better at estimating $p(D|H)$ than at estimating $p(H|D)$. We have noted in preceding sections some experimental evidence that tends to support this assumption. We have also noted some studies, however, that have shown that this result is not always found. Moreover, there is a question concerning how far the evidence that does support this assumption can be pushed. The only way that one can determine how accurately a man can estimate $p(D|H)$ is to observe his performance in experimental situations in which $p(D|H)$ is objectively defined or can be determined empirically. But, typically, in real-life situations of greatest interest, $p(D|H)$ is not known, and cannot be determined empirically--which is why is must be defined or can be determined empirically--which is why it must be estimated. The question arises then, if it is not known, how can we be sure that one's estimate of it is accurate? And the answer is that we cannot. How much confidence one should have in the conclusion that man is better at estimating $p(D|H)$ than at estimating $p(H|D)$ in real-world situations depends in large part on the extent to which one is willing to assume that what is known about performance in laboratory situations in which $p(D|H)$ usually has a straightforward relative-frequency interpretation is generalizable to real-world situations in which it does not.

Seventh, Bayes rule does not provide the decision maker with a criterion concerning when to stop processing incoming data and to make a decision. Inasmuch as data gathering can be costly in terms of both time and money, it is essential that any completely adequate prescriptive model of decision making have an explicit stopping rule to indicate when hypothesis evaluation should be terminated and a decision made.

We emphasize that these comments deal with limitations of Bayes rule. One might argue that the observations are unnecessary, on the grounds that proponents of Bayesian diagnosis have never claimed that these limitations do not exist. It seems to us important to make these limitations explicit, however, because they help to place the notion of Bayesian decision making in perspective. The idea of obtaining estimates of $p(D|H)$ or of likelihood ratios from humans and using these estimates to update posterior probability distributions in accordance with Bayes theorem is undoubtedly a reasonable approach to evaluation in some situations. It is not always appropriate or practicable, however, as some Bayesians have been careful to point out. Edwards (1967) describes the situations for which the approach is most appropriate as those that have one or more of the following three characteristics: "the input information is fallible, or the relation of input information to output diagnostic categories is ambiguous or uncertain, or the output is required to be in explicitly probabilistic form" (p. 71).

SECTION IX

PREFERENCE SPECIFICATION

It is generally assumed that a decision maker is not indifferent to which of the various possible decision outcomes occurs. As we have already noted, in some formal representations of decision situations, the decision maker's perferences with respect to the possible outcomes are made explicit in a payoff matrix. The contents of a cell of such a matrix is the worth to the decision maker of the choice of a specific action-alternative, given the truth of a specific hypothesis concerning the state of the world. The entire matrix presumably represents the situation fully: it identifies all the decision maker's action alternatives as well as all the possible states of the world, and shows for each alternative-state combination its worth to the decision maker.

9.1 A Difficult and Peculiarly Human Task

The problem is how to determine these worths. There are two observations to make in this regard. The first is that this task, more than any other associated with decision making, is peculiarly human. One would expect that many of the decision-related tasks that now must be performed by humans will in time be performed by computers. However, the specification of preferences for decision outcomes involves value judgments. To say that one decision outcome is better than, worth more than, or preferred to, another is to say that it represents a greater good within the context of the decision maker's own value system. Such judgments must come, at least indirectly, from man.

The second observation is that to specify one's preferences objectively is not necessarily an easy thing for an individual to do. Even when all of the action alternatives have been made explicit and the outcome of each possibility is known--that is, even when uncertainty is minimal--the decision task may still be a very difficult one. This is particularly true when the worths of the possible decision outcomes are intangible or depend on many factors. Consider, for example, the problem of choosing a house for purchase. Even assuming that one confines his attention to a few houses that he knows are available, and that he has all the information that he wants about each one, he has the problem of somehow deriving from many factors (purchase price, number of rooms, design, general condition, extras--porch, garage, storage space, extra baths, fireplace--lot location and layout, distance from work, tax rate in town, services and public facilities in town) a common figure of merit in terms of which one house can be judged to be more or less preferred than another.

In military situations, the specification of preferences may be especially difficult. It may often happen that none of the possible decision outcomes is intrinsically desirable, and the

decision maker may find himself faced with the necessity of attempting to choose the least undesirable one. The problem is aggravated by the fact that the assignment of preferences may necessitate the weighting of time, materiel, territory and human lives. One balks at the idea of trying to specify the value of human lives and that of a piece of territory in terms of a common metric, but this is what is done, at least implicitly, when a decision is made to attempt to gain a territorial objective when it is known that the endeavor is likely to result in the loss of a certain number of men. Or, consider the private transportation system in the United States. The builders, users and regulators of automobiles and highways have implicitly expressed a preference for a system that provides certain capabilities and conveniences at a cost of approximately 60,000 traffic fatalities per year. One suspects that the exercise of making explicit how the various factors that contribute to human preferences are traded off against each other in specific decision situations would often be revealing to decision makers themselves, who sometimes may have little conscious appreciation, without going through such an exercise, of how such factors do combine to determine their own preferences.

Among the eight aspects of decision making in terms of which this report is organized, preference specification is one of the two (the other is hypothesis evaluation) that have received the greatest amount of attention from philosophers and researchers alike. In the case of decision making under certainty, the study of preferences and the study of choice behavior amount to the same thing. Presumably one chooses what one prefers--and vice versa--if he can know for certain what the decision outcome will be.

9.2 Some Early Prescriptions for Choice

In order to make choices among alternatives that differ with respect to several incommensurate variables, one must, at least implicitly, derive from the several variables involved a single figure of merit with respect to which the alternatives can be compared. That is to say, one must be able to decide that in some global sense Alternative A is preferred to Alternative B. How this is generally done is not known; how it should be done is a matter of some dispute. Undoubtedly, individual methods for dealing with the problem range from highly intuitive impressionistic approaches (I just consider all the factors and decide that I like this combination better than that) to formal quantitative algorithms.

Benjamin Franklin was familiar with the problem, and his way of dealing with it is at least of historical interest: "I cannot, for want of sufficient premises, advise you what to determine, but if you please I will tell you how... My way is to divide half a sheet of paper by a line into two columns; writing over the one Pro, and over the other Con. Then, during three or four days' consideration, I put down under the different heads short hints of the

different motives, that at different times occur to me for or against the measure. When I have thus got them all together in one view, I endeavor to estimate the respective weights...[to] find at length where the balance lies... And, though the weight of reasons cannot be taken with the precision of algebraic quantities, yet, when each is thus considered, separately and comparatively, and the whole matter lies before me, I think I can judge better, and am less liable to make a rash step; and in fact I have found great advantage for this kind of equation, in what may be called moral or prudential algebra."*

A more formal attempt to procedurize choice behavior was made at about the same time by the British philosopher and social reformer, Jeremy Bentham. Starting with the basic premise that choices should be dictated by the extent to which their outcomes augment or diminish the happiness of the party or parties whose interest is in question (the "principle of utility"), Bentham attempted to define a quasi-quantitative procedure--a "hedonistic calculus"--the use of which would assure that the choices that are made would be consistent with this principle:

> "To take an exact account then of the general tendency of any act by which the interests of a community are affected proceed as follows. Begin with any one person of those whose interests seem most immediately to be affected by it, and take an account:
>
> (1) Of the value of each distinguishable pleasure which appears to be produced by it in the first instance.
>
> (2) Of the value of each pain which appears to be produced by it in the first instance.
>
> (3) Of the value of each pleasure which appears to be produced by it after the first. This constitutes the fecundity of the first pleasure and the impurity of the first pain.
>
> (4) Of the value of each pain which appears to be produced by it after the first. This constitutes the fecundity of the first pain, and the impurity of the first pleasure.
>
> (5) Sum up all the values of all the pleasures on the one side, and those of all the pains on the other. The balance, if it be on the side of pleasure, will give the good tendency

---

*This account of Franklin's approach to decision making was quoted by Dawes and Corrigan (1974), who found it in a letter from Franklin to his friend Joseph Priestly, dated September 19, 1772.

of the act upon the whole, with respect to the interests of that individual person; if on the side of pain, the bad tendency of it upon the whole.

(6) Take an account of the number of persons whose interests appear to be concerned, and repeat the above process with respect to each. Sum up the numbers expressive of the degrees of good tendency which the act has, with respect to each individual in regard to whom the tendency of it is good upon the whole; do this again with respect to each individual in regard to whom the tendency of it is bad upon the whole. Take the balance; which, if on the side of pleasure, will give the general good tendency of the act, with respect to the total number or community of individuals concerned; if on the side of pain, the general evil tendency, with respect to the same community" (Bentham, 1939, p. 804).

The value of a pleasure or pain, Bentham assumed, would depend on four factors:

" (1) Its intensity.
(2) Its duration.
(3) Its certainty or uncertainty.
(4) Its propinquity or remoteness."

Bentham did not expect that the procedure he defined would be "strictly pursued previously to every moral judgment, or to every legislative or judicial operation"; but he did contend that it represented a model of how judgments should be made, and a standard against which whatever procedures are used might be evaluated.

Bentham's approach to choice behavior can be, and has been, criticized on philosophical grounds. The principle of "the greatest pleasure for the greatest number" is itself open to criticism, because it appears to place no limits on the extent to which the many can prosper at the expense of the few, provided only that the "bottom line" of the calculation of the net happiness is increased in the process. For our purposes, the important point is the fact that Bentham attempted to reduce the process of making choices to a stepwise procedure.

9.3 Simple Models of Worth Composition

Although he used language that suggested that he believed that worth could be quantified and his procedure formalized as a sort of calculus for computing the worth of any given decision outcome, Bentham did not himself express his notions in mathematical form. His conceptualization of the choice process, however, is clearly suggestive of a linear model which expresses the worth of a decision alternative as a function of the sum of the

values of the various components of pleasure (or pain) that that alternative represents, weighted by the number of people that would be affected by the decision outcome. The choice would, of course, be the alternative with the greatest calculated worth.

Implicit in Bentham's prescription is the assumption that the total worth of a decision outcome is a monotonically increasing function of each of the factors which contribute to the worth, and that the monotone character of this relationship for any given factor is independent of the values of the other factors. Yntema and Torgerson (1961) have suggested that there are probably many practical choice situations in which this is a valid assumption. For example, the worth of a vocational choice probably increases monotonically with the attractiveness to the individual of the work involved, whatever the status of the other factors to be considered. Yntema and Torgerson present some data that suggest that when this is the case, the decision maker's choice behavior can often be matched, if not improved upon, by a selection algorithm that takes account only of how worth relates to each of the factors individually and ignores the ways in which the factors interact. To develop such an algorithm it is necessary only to determine how worth varies with the individual factors. Several ways of making this determination are suggested. An important point for our purposes is that the relationships of interest may be inferred from the behavior of the decision maker when confronted with the task of choosing between pairs of hypothetical alternatives selected to represent specific (in particular, extreme) combinations of the relevant factors.

Dawes and Corrigan (1974) have recently taken an even stronger position with respect to the practicality and the validity of simple linear decision algorithms in a wide variety of choice situations. They have shown that if each of the factors contributing to the worth of a decision outcome has a conditionally monotone* relationship to that worth, and the measurement of these factors is subject to error, then not only are decisions that are based on weighted linear combinations of the factors likely to be better than those made by human decision makers, but in some cases this is true even if the weights are equal or randomly chosen. Data from several studies of judgmental and choice behavior are reviewed in support of this conclusion. Of the situations reviewed by Dawes and Corrigan, the only ones in which a linear weighting algorithm did more poorly than a human decision maker were those in which the human's judgment was based on information not taken into account by the algorithm.

---

*A conditionally monotone relationship is one that is monotone, or can be made monotone by a scaling transformation.

9.4 The Problem of Identifying Worth Components

The implication is that if one can identify the factors in terms of which worth is determined, one frequently can improve significantly upon human judgment by application of a simple linear model. The problem, according to this view, is not in the development of arcane mathematical decision algorithms, or even in the application of complex weighting functions to a linear combination rule, but that of identifying the dimensions of the choice space and of determining how these dimensions relate, individually, to the worth of the possible decision outcomes.

The danger in this line of reasoning is that of assuming that identification of the factors in terms of which judgments are, or should be, made is a trivial task. As we have already suggested, such an assumption is almost certainly false for many, if not most, real-life decision situations. Most people can probably recall choices that they have made which they realize in retrospect were made without consideration of some factor that they would have recognized as relevant and important if only they had thought to think of it. An individual buys a house, for example, and realizes too late that he failed to determine whether the cellar leaks. Had the question occurred to him, he would have recognized it not only as a relevant consideration but as one that would have figured heavily in his assessment of the relative worths of candidate purchases. A potentially important aid to a decision maker would be a procedure that would facilitate the identification of the dimensions of his choice space. Having determined the factors upon which the relevant worths of possible choices depend, and how these factors relate functionally to worth, a simple linear model of the type espoused by Yntema and Torgerson might then be used to infer the decision maker's behavior in a choice situation. The experimental results reviewed by Dawes and Corrigan suggest that such a model might even be used in place of the decision maker to effect the choice.

9.5 Studies of Choice Behavior

In using the choices of a human as the standard against which to compare the performance of a model, one is assuming that humans behave in at least a consistent, if not an optimal, fashion. Only recently has the assumption that decision makers are able to make consistent choices among alternatives that differ on many dimensions without recourse to formal analytical procedures been tested.

Slovic and Lichtenstein (1971) have reviewed several approaches that have been taken to the problem of describing how people do in fact make such choices. They divide these approaches into two major categories: those that make use of correlational or regression analysis or the closely related analysis of variance, and those

that make use of Bayes theorem. Among the nonBayesian approaches that are reviewed are the correlation model of Hoffman (1960; 1968), the lens model of Brunswik (1952, 1956), the integration theory of Anderson (1968, 1969), and the theory of conjoint measurement of Luce and Tukey (1964) and Krantz and Tversky (1971). The objective in all of this work is to discover and describe how a human "judge" combines information concerning different attributes of a choice alternative to arrive at a judgment of its overall desirability relative to the other alternatives among which a choice is to be made.

The results of many of the studies reviewed by Slovic and Lichtenstein (1971) suggest that, although people can make "wholistic evaluations" (Fischer, 1972), they tend to focus their considerations on less than the full set of dimensions, and, as a consequence, frequently ignore potentially important information. Also, there appears to be a degree of random error in the evaluation process which increases as the decision maker attempts to consider increasing numbers of relevant attributes (Hayes, 1964; Kanarick, Huntington, & Petersen, 1969; Rigney & Debow, 1966).

On the basis of results obtained in his study of job-seeking behavior, Soelberg (1967) challenged the idea that people generally do make choices in accordance with worth-calculation models in real-world situations. In his words, "The decision maker believes _a priori_ that he will make his decision by weighting all relevant factors with respect to each alternative, and then 'add up numbers' in order to identify the best one. In fact, he does not generally do this, and if he does, it is done _after_ he has made an 'implicit' selection among alternatives" (p. 28). Soelberg draws a number of other conclusions from his study which, in the aggregate, seem to suggest that much of the effort that goes into decision making is calculated to rationalize--rather than arrive at--a choice. It's as though the decision maker were in cahoots with himself to deceive himself into perceiving his choices as well-founded when in fact the real basis for them may be unknown.

9.6 Procedures for Specifying Worth

Obviously people can--people do--make choices among multidimensional stimuli; the results mentioned above suggest, however, that our ability to handle many dimensions simultaneously in a consistent and reliable way without the aid of a formal procedure is somewhat limited. Given that the problem seems to be one of exceeding man's ability to process information, it is not surprising that some of the solutions that have been proposed take the form of ways of restructuring unmanageable problems so as to make them into problems of simpler proportions. Such procedures are sometimes referred to as decomposition procedures because they divide the task into subtasks that presumably are within the

decision maker's information-processing capabilities. The solutions to the subtasks are then used as a basis for inducing a solution to the original problem.

These procedures typically involve a number of steps (e.g., Fischer, 1972) such as specifying the alternatives to be compared, specifying the dimensions or factors with respect to which the alternatives are to be compared, assessing the worth of each alternative with respect to each dimension, and combining the results of the dimension-by-dimension assessments into some overall indicant of worth for each alternative. The first of these steps has not been a focus of attention in studies of preference specification; the alternatives usually are provided. In the real world, identifying these alternatives can be a nontrivial problem, but it is perhaps better thought of as a problem of information gathering than one of specifying preferences. The second step also has not received much research attention.

A great deal of attention has been given to the third of the steps mentioned by Fischer (e.g., Beck McClintock, 1967; Coombs, 1967; Fischer & Peterson, 1972; Fishburn, 1967; Hammond, 1967; Huber, Sahney, & Ford, 1969; Luce & Tukey, 1964; MacCrimmon, 1968; Miller, Kaplan, & Edwards, 1967; Raiffa, 1968). Numerous techniques have been proposed and studied for assessing the worths of alternatives with respect to individual dimensions or factors. These techniques range from simple, qualitative pair-comparison procedures that yield ordinally scaled preferences to relatively complex methods for deriving ratio scales for interdependent factors.

MacCrimmon (1968) has reviewed several prescriptive techniques for choosing among alternatives that differ with respect to multiple factors. The techniques that he considers are discussed under the following rubrics: (1) dominance, (2) satisficing, (3) maximin, (4) maximax, (5) lexicography, (6) additive weighting, (7) effectiveness index, (8) utility theory, (9) tradeoffs, and (10) nonmetric scaling. In each case, he describes the necessary assumptions and information requirements, and presents a formal mathematical representation of the optimal (or best) choice defined by the technique. Consideration is also given to the possibility of using several methods in combination on a given choice problem, as suggested earlier by Pinkel (1967).

A more recent review of worth-assessment techniques has been prepared by Kneppreth, Gustafson, Leifer, and Johnson (1974). In this review, methods are classified in terms of five properties: (1) whether probabilities are used, (2) what kind of judgment is required (e.g., simple preference, numerical assignment), (3) number of factors involved in a single judgment, (4) whether appropriate for continuous or discrete factors, and (5) nature of output produced (e.g., ranking of worth, quantitative indicant of worth).

Especially helpful features of this review are explicit discussions of what the authors see as the primary advantages and disadvantages associated with each of the methods described, and the provision of references for the theoretical bases of these techniques. Of particular relevance to this report is the stress that Kneppreth, et al. put on the need for training before some of these techniques can be used effectively.

The fourth step mentioned by Fischer--that of combining the results of factor-by-factor assessments into overall worth estimates--has proven not to be a difficult one in many practical situations because of the fact that a simple linear combination rule seems to work remarkably well in so many cases (see Section 9.3).

Prescriptive techniques for preference specification, or worth assessment, are of considerable interest because of the potential that they represent for procedurizing--and thereby, hopefully, simplifying--the solutions for complex choice problems. A less tangible but perhaps no less important benefit that can result from attempts to apply such prescriptive techniques in real-world situations stems from the fact that these procedures force the decision maker to be explicit concerning his own value system as it relates to the problem at hand. This fact has obvious ramifications vis-a-vis the problem of evaluating the performance of decision makers who make choices that affect the lives of others; one clearly wants to know, in such cases, not only what the choices are, but the bases on which they are made. Being forced to be explicit concerning the factors that determine his choice and the relative importance that he attaches to each of them may be as revealing to the decision maker himself as to an independent observer.

9.7 Preferences among Gambles

So far, we have considered only the problem of specifying preferences among stimuli that differ perhaps in many, but in known, ways. In this case the decision maker knows what the effect of any choice that he may make will be. Another type of preference specification that has been studied involves preferences among gambles, or between gambles and "sure things." The general procedure in such studies is to present the decision maker with a choice, either between two wagers, or, more typically, between a wager and a sure thing, and then to adjust either the possible outcomes of the wager(s) or the probabilities of these outcomes until the decision maker is indifferent to the alternatives from which he must choose. By repeating this process a number of times with different wagers, one can generate the kind of data from which worth functions can be inferred.

Typically, the wagers that have been used in these studies are such that one of the possible outcomes is more desirable than the other, and the probability of the less desirable outcome is unity minus the probability of the more desirable one. Slovic, (1967, 1969), however, has studied preference behavior in so-called duplex gambles in which the probabilities of "winning" or "losing" can be varied independently of respective payoffs. In this situation, the decision maker can win and not lose, lose and not win, win and lose, or neither win nor lose. As Slovic points out, "It can be argued that this type of gamble is as faithful an abstraction of real-life decision situations as its more commonly studied counterpart in which the probability of losing is equal to unity minus the probability of winning ($p_L = 1-p_W$). For example, the choice of a particular job might offer some probability (p_W) of a promotion and some probability (p_L) of a transfer to an undesirable location, and it is possible that one of these events, both of them, or neither of them, will occur" (p. 223).

In the first of Slovic's studies, two different methods of indicating the attractiveness/unattractiveness of a wager were explored. One method required the subjects to rate strength of preference directly on a scale ranging from +5 (strong preference for playing) to -5 (strong preference for not playing). The second required the subject to equate the attractiveness of this gamble with an amount of money such that he would be indifferent to playing the gamble or receiving the stated amount. One third of the subjects assigned to the second method were required to state the largest amount they would be willing to pay the experimenter in order to play each bet, and, for an undesirable bet, the smallest amount the experimenter would have to pay them before they would play it. Another third of the subjects were given ownership of a ticket for each gamble and required to state the least amount of money for which they would sell the ticket. The subjects in the final third were required to state a fair price for a given gamble in the absence of information as to whether they or the experimenter owned the right to play it.

Slovic demonstrated that subjects did not weight the risk dimensions in the same way when bidding as when rating. Variation in the ratings was influenced primarily by variation in probability of winning (p_W), while variation in bidding was influenced primarily by variation in probability of losing (p_L). Also, payoff dimensions--dollars won ($W) and dollars lost ($L) produced more effect on bids than on ratings, while probability dimensions produced more effect on ratings than on bids. Finally, it was found that when a person in the bidding group considered a bet to be attractive, his judgment of its degree of attractiveness was determined primarily by the amount ($W); when he disliked a bet, the primary determinant of the degree of dislike was ($L). This finding has particularly important methodological implications, because, as Slovic points out, no existing prescriptive theory of decision

making would consider that response mode should be a determinant of the way in which decision makers utilize probabilities and payoffs in making decisions under risk, and he argues that behavior in such circumstances may be strongly influenced by information-processing considerations.

9.8 Preference Specification and Training

On first thought, preference specification--among all the tasks associated with decision making--might appear to pose the least challenge for training research. One might assume that if there is an aspect of decision making that comes naturally, it should be that of saying what one's preferences are. Things clearly are not that simple, however, and the evidence is abundant that people do not always know what their preferences are, or at least how to specify them in an unambiguous and consistent way.

The research reviewed in this report suggests at least four problems that relate to training and preference specification. First is the question of how to train people to make judgments of subjective probability that are independent of the worths of possible decision outcomes, as the use of subjective expected utility models requires (see Section 2.2). A second and closely related question is that of how to train people to make worth judgments that are invariant across different measuring techniques.

The development of decomposition methods has been motivated by an interest in simplifying the process of making preferences, and their bases, explicit. As Kneppreth, Gustafson, Leifer, and Johnson (1974) have pointed out, however, some of these procedures, particularly those that yield the most quantitative results, are workable only with relatively sophisticated users. A third challenge for training research, therefore, is to develop methods for providing the necessary training in cost-effective ways.

A fourth problem relates to two aspects of decision making, preference specification and information gathering. In laboratory studies of choice, the dimensions in terms of which preferences are to be specified typically are given. In real-world situations, however, the dimensions of choice are often determined by the decision maker himself; in other words, the factors that are considered in attempting to assess the relative merits of the choice alternatives are those that the decision maker happens to think about. Surprisingly little attention has been given by researchers to the question of how capable people are at enumerating on demand the factors that they would consider important in any particular choice situation. It is not even clear whether, when provided with a list of such factors, one can say with confidence whether the list is complete. Much more research is needed, both to determine

human limitations and performance characteristics in this regard, and to explore how training might improve one's ability to make one's worth space explicit vis-a-vis specific choice problems.

SECTION X

ACTION SELECTION

Selection, or choice, is often thought of as representing the essence of decision making. And obviously, if one has no options, then he has no decisions to make. Paradoxically, however, the act of choosing per se is the least interesting of the aspects of decision making that are considered in this report. This is because of the fact that when the other aspects have been realized--when information has been obtained, the decision space structured, hypotheses generated and evaluated, and preferences stated--the choice may, in effect, have been determined. This is, of course, as it should be. One's goal in all of these activities is to remove, insofar as possible, doubt about what the choice should be.

In spite of his best efforts to reduce uncertainty to a minimum, and thereby to discover what his decision ought to be, however, the decision maker may, on occasion, feel very much "left to his own devices" when forced to make a choice. Ellsberg (1961) rather graphically described the sense of frustration that one can feel when he faces his moment of truth and is not entirely convinced of the adequacy of the basis on which the choice will have to be made. "(This) judgment of the ambiguity of one's information of the overall credibility of one's composite estimates, of one's confidence in them, cannot be expressed in terms of relative likelihoods or events (if it could, it would simply affect the final, compound probabilities). Any scrap of evidence bearing on relative likelihood should already be represented in those estimates. But having exploited knowledge, guess, rumor, assumption, advice, to arrive at a final judgment that one event is more likely than another or that they are equally likely, one can still stand back from this process and ask: 'How much, in the end, is all this worth? How much do I really know about the problem? How firm a basis for choice, for appropriate decision and action, do I have?' The answer, 'I don't know very much, and I can't rely on that,' may sound rather familiar, even in connection with markedly unequal estimates of relative likelihood. If 'complete ignorance' is rare or non-existent, 'considerable' ignorance is surely not" (pp. 20,21).*

Most of the decision situations that we have considered in this report involve the problem of choosing one from among several

---

*This statement is contained within a larger discussion of circumstances in which it may be "sensible" to act in conflict with the prescription of the Savage (1954) axioms (see Section 2.2). The reader is referred to the full discussion for an interesting analysis of the problem of ambiguity in choice behavior.

courses of action. It is important to note, however, that people sometimes find themselves faced with the task of deciding not what to do, but when to do it. The required action may be dictated by circumstances, or predetermined in one way or another, but the individual is left with the job of deciding on the best time to act. This type of decision problem is nicely illustrated by the following situation.

Consider a pistol duel in which the duelists are instructed to turn to face each other on signal and to fire one shot at will. Suppose that once the men have faced each other, each may walk toward the other, reducing the distance between them if he wishes. We may assume that the accuracy of each duelist improves, although not necessarily at the same rate, as the distance between them decreases. Clearly, each man faces a dilemma: every second that he delays firing in order to decrease the distance between him and his opponent and to increase his chances of an accurate shot, he also increases the chances of success for his opponent; on the other hand, if he fires too soon, he risks missing, in which case his opponent is free to advance on him until his shot will be certain to find its mark.

This type of situation is representative of what Sidorsky, Houseman, and Ferguson (1964) have characterized as "implementation-type decision tasks." In Sidorsky's experiments the duelists were simulated navy tactical units, but the problem was essentially the same as that of the individual antagonists. The decision maker had to decide when to fire a missile, knowing that both the probability of hitting his opponent and the probability of being hit by him were increasing (but at different rates) in time.

A particularly interesting result from this work is the finding that subjects performed less appropriately when operating at a disadvantage than when operating at an advantage. One of the conclusions that Sidorsky and his colleagues drew from the results of a series of studies (Sidorsky & Houseman, 1966; Sidorsky, Houseman, & Ferguson, 1964; Sidorsky & Simoneau, 1970) was that "the inability to analyze and respond appropriately in disadvantageous situations is a major cause of poor performance in tactical decision making" (Sidorsky & Simoneau, 1970, p. 57). If this observation is generally valid, its implications for tactical decision making are clearly very significant. The implications for training are also apparent, namely, the need for extensive decision-making experience in disadvantageous situations.

SECTION XI

DECISION EVALUATION

The problem of evaluating the performance of decision makers is a difficult one and it is critically important to the task of training. Without an evaluation scheme, there is no way of ascertaining whether training has resulted in an improvement in decision-making performance. Training assessment is not the only reason for an interest in evaluation of decision-making performance, however. Anyone who finds himself in a position of having to pass judgment on the performance of a decision maker is in need of a set of criteria in terms of which that judgment can be made. Moreover, a decision maker himself might wish to evaluate a particular decision that he has made in terms of a set of objective criteria.

Unfortunately, a completely satisfactory set of objective criteria against which performance can be compared has not been developed. As Kanarick (1969) has pointed out, "unlike other behaviors, there is no standard dependent variable, such as time-on-target, trials to criterion, or percent correct." One can, of course, choose for study in the laboratory only tasks for which performance <u>can</u> be objectively evaluated (e.g., probability estimation for frequentistic events); however, one runs the risk of thereby excluding from study a large percentage of the problems of interest. Certainly, in most real-life decision situations in which the objectives are complex, the stakes are real, and the information is incomplete, evaluation is an extremely difficult task.

11.1 Effectiveness versus Logical Soundness

Of central importance to a discussion of evaluation of decision making is the distinction between effectiveness and logical soundness. Failure to make this distinction sharply--sometimes to make it at all--has resulted in much confusion in the literature. Effectiveness and logical soundness are quite different things. One might be willing to assume that logically sound decisions will, on the average, tend to be more effective than decisions that are not logically sound. However, the assumption that the correspondence will necessarily hold in any particular instance is manifestly not valid.

A decision is effective to the extent that the result to which it leads is one which the decision maker desires. Effectiveness usually is easily determined after the fact. The logical soundness of a decision depends on the extent to which the decision maker's choice of action is consistent with the information available to him at the time the decision was made, and with the decision maker's own preferences and goals. That these are quite d ferent factors is clear from a simple example. Suppose that

one is given the option of betting $5 against $20 that the next roll of a fair die will come up 6, or betting $10 against $12 that the up face on the next roll will have an odd number of dots. If he elects to make the first bet and the roll produces a 6, we would say that the decision was an effective one. However, whether it could be considered a logically sound one would depend on what the decision maker's objectives were. If his intent was to maximize his potential gain, or to minimize his potential loss, the decision was sound. If his intent was to maximize his expected gain, it was not.

Decision-making behavior should be evaluated in terms of its logical defensibility and not in terms of its effectiveness, inasmuch as effectiveness is found to be determined in part by factors beyond a decision maker's control, and usually beyond his knowledge as well.* It often appears not to work this way in practice, however. Evaluation of decisions in terms of their outcomes seems to be the rule, for example, in the world of finance and business. Investment counselors are hired and fired on the basis of the consequences of their portfolio recommendations, and corporate managements are frequently juggled as a result of unsatisfactory profit and loss statements. Although the cliché "it's the results that count" has particularly strong intuitive appeal in this context, decision outcome is no more justified as <u>the</u> basis for evaluation of decision making in the financial world than in any other. As Krolak (1971) asserts in a discussion of portfolio management evaluation: "The real question to be answered is how well did [I] do w[illegible] the information, capital, strategy and ability to assume risk[illegible] compared with others who might possess the same resources?" (p. 235).

That decision-making performance in military-training situations is <u>not</u> always evaluated in terms of its logicality, has been noted by Hammell and Mara (1970). In discussing some of the mission

---

*Commenting on Fuchida and Okumiya's account of the WWII Battle of Midway, Admiral Spruance (1955) made the following interesting observation: "In reading the account of what happened on 4 June, I am more than ever impressed with the part that good or bad fortune sometimes plays in tactical engagements. The authors give us credit, where no credit is due, for being able to choose the exact time for our attack on the Japanese carriers when they were at a great disadvantage--flight decks full of aircraft fueled, armed and ready to go. All that I can claim credit for, myself, is a very keen sense of the urgent need for surprise and a strong desire to hit the enemy carriers with our full strength as early as we could reach them."

training that is carried out in ASW tactical training programs, they point out that performance evaluation is based, in many instances, on the simple effectiveness indicator of whether or not the team scores a hit. If it does, performance is judged to be acceptable. Commenting on specific training exercises that they observed they note: "If a hit was made, regardless of circumstances, each team member's performance was usually considered good... In some instances a hit was scored because the target would make a predetermined maneuver into the path of a torpedo which had been obviously fired in a wrong direction" (p. 9).

It is probably safe to assume that most people in decision-making positions are more likely to be rewarded, or censured, as the case may be, on the basis of the effectiveness of their decisions than on that of their logical quality. This is due in part perhaps to the fact that society is far more interested in the results produced by its decision makers than in the reasons for which decisions were made. It is undoubtedly also true, however, that it is easier to determine the outcome of a decision than to determine whether the decision was logically justified at the time that it was taken. One wonders how many heroes have been made, not in spite of, but because of, very poor decisions which have had happy outcomes, and, conversely, how many "bumblers" owe their reputations not to the illogicality of critical decisions they have made, but to fortuitous turns of events that have blessed sound choices with disastrous results.

We may note in passing that even if one wishes to evaluate a decision in terms of its effectiveness, rather than its logical soundness, the task may be less than straightforward. Miller and Starr (1969) make the point that decision objectives are not always singular. Often, one is attempting to realize several objectives simultaneously, and seldom is it possible to optimize with respect to all objectives at the same time. It is difficult in such cases to evaluate a decision outcome unless its implications with respect to all the objectives can be combined into a single figure of merit.

One attempt to develop a procedure for combining performance scores on various decision-effectiveness criteria into a single figure of merit was made by Sidorsky (1972), and Sidorsky and his colleagues (1968, 1970). A set of operational criteria that were intended to be used to evaluate the decision performance of a military tactical unit was identified as follows: spatial relationships (the spatial interface between own and enemy tactical units), self-concealment (the degree of success in keeping the enemy uninformed concerning own unit), information generation (the degree of success in keeping informed concerning enemy unit), weapon utilization (destroy or counterattack capability), and conservation of resources (adequacy of supplied). Such criteria have been used by Sidorsky to rate the quality of

decisions made during experimental tactical exercises. A Decision Response Evaluation Matrix was developed which, when used in conjunction with an algorithm for combining scores with respect to all five operational criteria, permitted the quality of a decision to be expressed as a single measure.

11.2 Evaluation Criteria

Granted that logical soundness is the appropriate basis on which to evaluate decisions, the problem then is to translate that principle into a set of objective criteria against which decision-making performance can be judged. In view of the huge literature on decision making, surprisingly little attention has been given to this problem.

Sidorsky and his colleagues (1964, 1966, 1968, 1970) and Hammell and Mara (1970) have suggested five behavioral factors in terms of which an individual decision-maker's performance might be judged: stereotopy (the tendency of a decision maker to respond in an unnecessarily predictable way), perservation (the tendency to persist when persistence is unwarranted), timeliness (the extent to which the decision-maker's behavior is reasonable in terms of the time constraints imposed by the situation), completeness (the extent to which all available relevant information is used), and series consistency (the consistency of the decision-maker's behavior within the context of a series of interrelated actions). The first two factors are liabilities for a decision maker; the last three are assets. In contrast with the operational criteria mentioned in the preceding section, these behavioral criteria are more concerned with the logicality of a decision than with its effectiveness.

The conceptualization of the decision-making process that has provided the structure of this report suggests a number of dimensions with respect to which the quality of a decision-making activity might be evaluated: the adequacy of the information-gathering process; the sensitivity of data evaluation; the appropriateness of the structure that is given to a decision problem; the facility with which plausible hypotheses are generated; the optimality of hypothesis evaluation; the sufficiency with which preferences are specified; the completeness of the set of decision alternatives that is considered; the timeliness of action selection and its consistency with the decision maker's preferences, objectives, and information in hand. The development of techniques for assessing these aspects of decision making quantitatively and unambiguously represents a challenge to investigators of decision-making behavior.

11.3 A Methodological Problem

It is worth noting that to determine after a decision has been made whether its basis was logically sound may be a very difficult

task. People usually can give plausible reasons for choices they have made. One may be permitted a certain amount of skepticism, however, concerning whether reasons that are given after the fact are the reasons that prevailed at the time of the making of the choice (Soelberg (1967). This is not to sugg st that people necessarily misrepresent the bases for their decisions intentionally. It seems not unlikely, however, that we frequently convince ourselves, without being conscious of doing so, that choices have been determined by certain rational considerations, when in fact those considerations were discovered or invented only after the choice was made. One might argue that even though the alleged basis of a decision may not have been verbalized, or even consciously appreciated by the decision maker, it could still have been operative at a subconscious level at decision time. But this is a difficult, if not impossible, point to confirm or invalidate experimentally, and for that reason it is not a very useful hypothesis. Pascal (1910) expressed his skepticism concerning the credibility of after-the-fact introspective explanations of behavior over three hundred years ago: "M. de Roannex said: 'Reasons come to me afterwards, but at first a thing pleases or shocks me without my knowing the reason, and yet it shocks me for the reason which I only discover afterwards.' But I believe, not that it shocked him for the reasons which were found afterwards, but that these reasons were only found because it shocks him" (p. 98).

SECTION XII

SOME FURTHER COMMENTS ON TRAINING OF DECISION MAKERS

Throughout this report we have commented on how the theoretical notions and research findings that have been reviewed relate to issues of training and training research. These comments have been made within the contexts of the discussions to which they pertain. It is not our purpose in this section to review or summarize these comments, but rather to turn to some training-related topics that have not been addressed elsewhere in the report.

12.1 Performance Deficiencies versus Performance Limitations

Some investigators (Hammell & Mara, 1970) have advocated the approach of identifying "behavioral deficiencies" and developing training programs that are designed to ameliorate them. Similarly, Kanarick (1969) has suggested that one component of a training program for decision makers should be that of making them aware of some of the common reasons for the making of poor decisions.

The term "deficiencies" has been used in two ways in the literature: to refer to stereotyped ways of behaving suboptimally, and to refer to basic human limitations. In what follows, we will refer to the second type of "deficiencies" as limitations, and use the word deficiency only to denote suboptimal but presumably correctable behaviors. An example of a behavioral deficiency would be the tendency of humans to be overly conservative in their application of probabilistic information to the evaluation of hypotheses. A possible example of a limitation would be the inability of most people to weigh more than some small number of factors, without some procedural help, in arriving at a preference among choice alternatives.

The distinction between deficiencies and limitations has important implications for training. Deficiencies may be "trained out"; basic limitations must be "trained around."

The first problem in dealing with either a putative deficiency or a limitation, however, is to verify that it indeed exists. It is obviously imperative, when a deficiency or limitation is identified by a single experimental study, that the finding be corroborated by further research. More important, however, and more difficult, is the problem of establishing that the conclusions drawn from experimental studies are valid beyond the laboratory environments in which the results were obtained. It is exceedingly difficult to capture some of the aspects of many real-world decision problems (e.g., very high stakes) in laboratory situations. And what may constitute appropriate behavior in the one situation may prove to be inappropriate in the other.

Assuming, however, that one is able to identify some examples of deficient behavior that appear to be fairly universal among decision makers, the question is how to go about training them out. One obvious possibility is to expose trainees to decision-making situations in which a given deficiency is likely to show itself if it is ever going to do so, and then provide the individual with some immediate feedback concerning the appropriateness of his behavior. One would probably want to provide numerous opportunities for the same deficiency to show itself in a variety of contexts, providing feedback to the trainee each time that the deficiency is displayed. Probably, too, feedback should be provided for some time after performance has improved to the point that the deficiency is no longer apparent.

When dealing with basic human limitations, the goal should be to educate the decision maker concerning what those limitations are and to provide him with the means for working around them. For example, if it is the case that without the help of some explicit procedure, a decision maker cannot effectively weigh more than n variables in attempting to optimize his choice of an action alternative, it may be futile to try to train him to make effective use of more variables; however, if that is the case, he should be made aware of his limitation and be trained to perform within it.

Another approach to dealing with deficiencies and limitations--in addition to training--is that of providing the decision maker with aids to facilitate various aspects of the decision process. The goals of training and of decision aiding are not viewed by the writers as mutually exclusive, but rather as complementary, approaches to the improvement of decision making. Moreover, the fact that decision aids are being developed has implications for training, a point to which we will return in Section XIII.

12.2 Simulation as an Approach to Training

A common approach to the problem of training decision makers is that of simulation (Bellman, Clark, Malcolm, Craft, & Ricciardi, 1957; Cohen & Rehman, 1961). The idea is to place the decision maker in contrived situations that are similar in certain critical respects with the decision-making situations that they are likely to encounter in the real world. The approach has been used in efforts to train business executives (Martin, 1959), prospective high-school principals (Alexander, 1967), research and development project managers (Dillman & Cook, 1969), military strategists and tacticians (Carr, Pyrwes, Bursky, Linzen, & Hull, 1970; Paxson, 1963), high-school history and science teachers (Abt, 1970), vocational-education leaders (Rice & Meckley, 1970), and government planners (Abt, 1970).

Most business colleges and graduate schools today make some use of simulation and gaming techniques to teach management and decision-making skills. Also, as a result of early efforts by the American Management Association to develop a decision-making course, corporations such as General Electric, Pillsbury, Westinghouse, and Standard Oil of New Jersey have devised in-house training programs that make use of simulation techniques.

Two different forms of management-training games are discussed by Cohen and Rhenman (1961) in their survey of the present and future roles of such games in education and research. The first form--the "general-management" game--attempts to provide experience in the making of business decisions at a top-executive level, while the second form--the "functional" business game--focuses on specific decision situations within a limited functional level of the organization. Because of the complexity of interactions among organizational entities and the multidimensionality of the decision environment simulated in the general-management games, the possibility of defining and utilizing optimal strategies has not yet been demonstrated. The functional game situations, on the other hand, which are typically lower in complexity, allow for the specification and application of optimal or "best" strategies.

A variety of views have been expressed concerning the strengths and weaknesses of simulation as an approach to training. Kibbee (1959) suggests the following advantages:

> "1) It (simulation) can provide a dynamic opportunity for learning such management skills as organization, planning, control, appraisal, and communication.
>
> 2) Simulation can provide an executive with an appreciation of overall company operations and the interaction between man, money and materials. It helps make a generalist out of a specialist who has never had the opportunity of reviewing his decisions as they affect the organization as a whole.
>
> 3) Simulation can provide executives with practice, insight and improvement of their main function: making decisions. Faced with realistic decisions about typical business problems, they can experience years of business activity in a matter of hours, in an environment similar to that they face in everyday life.
>
> 4) Simulation can exhibit what Dr. Forrester of M.I.T. calls the 'dynamic, ever-changing forces which shape the destiny of a company.' The general business principles that are illustrated can be studied and understood by the participants" (p. 8).

Similar themes are expressed by Abt (1970) concerning the efficacy of management games:

> "Games are effective teaching and training devices for students of all ages and in many situations because they are highly motivating, and because they communicate very efficiently the concepts and facts of many subjects. They create dramatic representatives of the real problem being studied. The players assume realistic roles, face problems, formulate strategies, make decisions, and get fast feedback on the consequences of their action.
>
> In short, serious games offer us a rich field for a risk-free, active exploration of serious intellectual and social problems" (p. 13).

Simulation, as a general approach to training of decision makers is not without its critics, however. Martin (1959), who generally endorses the approach, volunteers several caveats. He points out, for example, that many of the qualitative dimensions of a situation, such as personnel quality and morale in an organization being modelled, are difficult to reflect in a game. Further, in order to make a game administratively manageable, it may be necessary to limit the degrees of freedom one has with respect to innovation, which is an unfortunate constraint. Finally, he points out that it is not always clear exactly what students are learning in a simulation situation. "There is no doubt that the simulation technique is a powerful teaching device, and therefore is potentially dangerous unless we are relatively sure of what is being taught."

One wonders, in connection with the last point, if definition of what should be taught and learned can really be expected prior to development of an adequate prescriptive theory of management decision making. Moreover, it seems clear that so long as decisions are evaluated in terms of effectiveness rather than in terms of logical soundness, the answer to the question of whether any training program is teaching individuals to make optimal decisions will remain a matter of conjecture. Apropos the point of how to insure that simulations have some realism, Freedy, May, Weisbrod, and Weltman (1974) have proposed a technique for generating decision-task scenarios that utilize expert judgments concerning state variables and transformations in much the same way that a Bayesian aggregator would make use of expert judgments of conditional probabilities.

We would summarize our own attitude toward simulation training in the following way. The approach has many advantages. The student can be exposed to a variety of decision situations. Situation parameters can be varied systematically, thus permitting the

study of their effects on decision-making performance. The consequences of incorrect decisions are not catastrophic, as they could be in some real-life situations of interest. The student's performance can be evaluated and immediate feedback can be provided to him, thus, presumably, improving his chances of learning.

On the negative side of the ledger, there is first the difficulty of the task of deciding what aspects of a situation to simulate. Any simulation is a simplification, and if one wishes to assure transfer of what is learned in the simulated situation to real-life situations, it is imperative that the simulation preserve those aspects of the real-life situation that are relevant to the skill that is being trained. Moreover, the difficulty of assuring the veridicality of a simulation is likely to increase greatly with the complexity of the situation that is being simulated. Second, there is the problem of generality. Situations are specific. One wants the student to carry away from training sessions skills which will be applicable in a variety of contexts. Simulation itself does not guarantee that that will occur. In fact, one might guess that there would be the danger of focusing on specific aspects of particular situations which could have a tendency to impair the learning of general principles.

12.3 On the Idea of a General-Purpose Training System for Decision Makers

A training system for decision makers that has a reasonable degree of generality is bound to be a relatively complex system. Moreover, given the current level of understanding of decision processes, it is unlikely that anyone would be able to design a system that would be certain to be satisfactory. The approach that seems to us most likely to produce a useful system is an explicitly evolutionary one, and one that involves potential users of the system in its development from the earliest stages. What one needs to do is build a working system that represents one's best guess concerning what capabilities such a system should have, and then elaborate, extend, and improve the system in accordance with the insights that are gained through attempts to make use of it.

The idea that many complex systems are best developed through an evolutionary process is not a new one. Benington (1964) has argued strongly for such an approach in the development of command-and-control systems. Commenting on the fact that many systems become obsolete even before they are operational, he notes that "The principal cause of this situation is the fact that until recently the proposed users of these systems did not take many interim steps that would have helped them; instead, they waited for the grand solution. When the development of these command-and-control systems was undertaken, it was thought that the design team could analyze present operations, project changes over many years, design a system for the far-off future, and then implement. Now most agree that this process just won't work" (p. 16).

SECTION XIII

DECISION AIDS

The recognition that--whether because of behavioral deficiencies or basic limitations--men often do not perform optimally as decision makers has motivated the development of numerous decision-aiding procedures and techniques. The existence of decision aids has two somewhat opposing implications for the training of decision makers: On the one hand, insofar as an aid succeeds in simplifying or otherwise facilitating the performance of some specific task, its existence may lessen the training demands vis-a-vis that task; on the other hand, users of decision aids must be trained to use those aids. It does not follow from the fact that some training may be required before an aid can be used effectively that the aid is therefore a failure; if a trained user of an aid can make better decisions than a trained decision maker who does not use that aid, then the aid may be said to be an effective one.

Given the view of decision making as comprised of a variety of tasks and processes, it seems reasonable to expect that initial decision-aiding techniques will be more successfully applied to some of these tasks than to others. The goal should be, not to develop <u>the</u> grand aid for <u>the</u> decision maker, but, rather, to develop <u>a</u> variety of aids <u>to</u> facilitate performance of the various tasks. Together, a group of such aids might be thought of as a "decision support system" (Levit, Alden, Erickson, & Heaton, 1974; Meadow & Ness, 1973; Morton, 1973), but the individual aids, and not the system, are probably the more reasonable objectives toward which to work initially.

Another factor that some researchers have argued is highly relevant to the design of decision aids is that of individual differences. One group of investigators, for example, has characterized "decision styles" in terms of three dimensions with respect to which individuals are assumed to vary: abstract-concrete, logical-intuitive, active-passive (Henke, Alden, & Levit, 1972; Levit, Alden, Erickson, & Heaton, 1974). All possible combinations of the extremes of these dimensions are viewed as eight "pure decision styles" that are representative of the types of individualized approaches to decision making that decision-aiding systems must take into account. The point that these investigators make is that decision aids or decision support complexes, should be designed with particular users, or user types, in mind. Systems designed for one type of decision style, they claim, may degrade the performance of a user who operates according to a different style.

Decision aids run the gamut from the types of heuristic principles discussed by Polya (1957) to explicit paper and pencil procedures for working through some aspect of a decision problem, to interactive computer-based techniques. In this section, we consider only a few of the many aids to decision making that have been developed. The intent is not to provide an exhaustive review but a representative sampling of what has been done in this regard.

13.1 Linear Programming

Linear programming is a mathematical technique for determining a set of decision parameter values that maximizes or minimizes specified functions within certain linear constraints. The technique is particularly useful in solving such problems as resource allocation, production mix and industrial cost control. It is best illustrated by a simple example.

Suppose a manufacturer produces three products. We will designate the monthly quantities of these products as x_1, x_2 and x_3. The products have different unit production costs, say, a_1, a_2, and a_3, and different unit sale prices, say, b_1, b_2, and b_3. To keep the illustration simple, we ignore the problem of inventories. Raw material limitations restrict the number of units of products 1 and 3 that can be produced per month to c_1 and c_3, respectively. The total number of man-hours available to the producer is n per month, and it requires d_1, d_2, and d_3 man-hours to produce one unit of products 1, 2, and 3, respectively. The problem is to determine the number of units of each product that the manufacturer should produce per month in order to maximize his profit.

Linear programming is a technique for solving such problems, when solutions exist. The technique involves expressing the constraints as a set of simultaneous linear equations, and then

searching within the ranges of the values of the independent variables that satisfy the equations for those values that optimize the desired function. In the case of our example, the function to be optimized (in this case, maximized) would be the profit function, i.e.,

$$(b_1-a_1)x_1+ (b_2-a_2)x_2 + (b_3-a_3)x_3.$$

When the problem involves only two or three decision variables, a geometrical model of the situation can give the decision maker an intuitively meaningful representation of the significance of the various factors and, in particular, of the sensitivity of the decision outcome to a less than optimal selection of values for the decision variables. When certain boundary conditions are met, the set of parameter values that satisfies the linear constraints within which the decision must be made is represented by convex polygons or polyhedra (in the two- and three-variable cases, respectively), and the solution to the optimization problem invariably is (or at least contains) one of the figure's vertices. The same principle holds in cases of more than three variables, but, of course, the geometrical model is no longer helpful.

One of the limitations of linear programming is the fact that it is applicable only to situations in which the decision space has been fully represented numerically and the outcomes of all of the admissible decisions are known. Another is the fact that it can be used only when the effects of the individual decision variables combine in an additive (linear) fashion. One can imagine real-life decision situations in which the effect of a change in the value of one decision variable depends in some way on the value of another variable. For example, how much importance one would attach to a difference in salary between two jobs might depend on whether the jobs also differed significantly in terms of the extent to which they placed one's life in danger. As has already been noted in Section IX of this report, however, several investigators of decision making have argued that the assumption of additivity appears to be a reasonable one in many, if not most, real-life situations. Probably the more difficult requirement to satisfy in order to use linear programming techniques is that of adequately structuring the decision space and quantifying the salient variables. When the necessary conditions can be met, however, there can be no doubt of the effectiveness of the technique.

13.2 Decision Trees and Flow Diagrams

Sometimes it is possible to convert an apparently complex set of written or verbal instructions concerning a problem-solving procedure into a decision tree or flow diagram. When such a conversion can be accomplished, it is often found that the desired

procedure is more easily and efficiently followed with the aid of the diagram than with the original set of instructions (Blaines, 1973; Raiffa, 1968; Wason, 1968; Wright, 1971).

The following distinction between decision trees and flow diagrams is made by Triggs (1973): "A decision tree is an assembly of individual paths in a structure organized so that no path ever returns or proceeds to another part of the diagram. A decision flow diagram may, on the other hand, contain paths that return to early parts of the diagram or feed to other common elements. A decision flow diagram can be more operationally directive in its structure, and less concerned with the explicit details of the decision process. In a tree structure, at every node of the tree, the user of the diagram can exactly state by what set of chance events and decisions one arrived there. The flow diagram structure is not always organized so that each such path can be uniquely specified" (p. 3).

The clarity and efficiency gained by representing procedures requiring sequential decisions in diagrammatic form have been recognized for some time. In such fields as computer programming and systems analysis, graphic techniques have been employed in the teaching and conduct of specific programming, debugging, maintenance, and troubleshooting tasks. Only recently, however, have formal attempts been made to assess the benefits to be derived. In an entertaining article by Davies (1970) the results of a relevant experiment by B. N. Lewis are discussed. The latter investigator presented a series of six problems involving a tax regulation to each of 60 subjects. One third of the subjects worked with the original (prose) statement of the regulation, a second third worked with a simplified (prose) statement, and the final third worked with an algorithmic (decision tree) form. The mean time required by the original prose group to solve all six problems was 23.4 minutes, compared to 11.8 minutes required by the simplified prose group and 9.2 minutes required by the algorithm group. Mean errors in problem solution followed a similar pattern: 29%, 10%, and 8% for the respective groups.

More recently, Blaiwes (1973) compared the performance of decision makers who had been given instructions concerning the construction and use of decision trees with that of decision makers who had not been so instructed. Only one of the ten subjects in the uninstructed group gave evidence of using a decision-tree approach to the solution of the four experimental tasks, whereas all ten of the instructed subjects used it. Subjects using the decision-tree approach initially required more time than uninstructed subjects, but their performance improved as they gained facility with the approach. Most importantly, subjects in the instructed group performed at a higher level of

accuracy than subjects in the uninstructed group. Although the possible effects due to practice cannot be separated from those due to problem difficulty because of the particular design used by Blaines, we regard the experiment as a demonstration of the ease with which the decision-tree approach can be taught to individuals who have not previously encountered it.

A review of numerous attempts to apply decision trees and flow diagrams to the solution of decision problems (e.g., Baker, 1967; Clarkson, 1963; Dutton & Starbuck, 1971; Horabin, 1972; Howard, Matheson, & North, 1972; Rousseau & Zamora, 1972; Tuddenham, 1968) has been prepared by Triggs (1973). He points out that the degree to which such aids can be useful to a decision maker will depend on the nature of the problem that is faced. They tend to be most useful for situations that are easily structured, perhaps by means of decomposition techniques advocated by Raiffa (1968). Triggs cautions against the temptation "to make a complex problem tractable by forcing it into a conceptual representation with which one knows how to cope," at the expense of ignoring or eliminating critical aspects of the real problem. He also points out that the task of imposing the type of structure on a decision problem that is necessary if decision trees or flow diagrams are to be used to advantage, may be sufficiently time-consuming and expensive to assure its impracticality in some dynamic situations in which the time for analysis is limited. Moreover, forcing the decision maker to think about his problem in terms of a specific structure may inhibit his use of cognitive skills that he otherwise might bring to the task. Triggs concludes, however, that on balance these cautions do not negate the efficacy of the approach. Citing Zadeh's (1973) work, he notes that "even in systems that are too complex or too ill-defined to admit of precise quantitative analysis, 'fuzzy' algorithms and diagrams have the potential of being useful to the human decision maker" (p. 17).

A lucid tutorial treatment of decision trees and their use is presented by Peterson, Kelly, Barclay, Hazard, and Brown (1973) in Chapters 2 and 3 of a Handbook for Decision Analysis. The handbook has been prepared for the express purpose of aiding the individual who is faced with substantive decision problems to apply concepts and procedures of decision theory to the solution of those problems.

13.3 Delphi, an Aid to Group Decision Making

The decision maker of most prescriptive models of decision making could be an individual, a committee, a corporation, or a machine, inasmuch as such models are concerned with the decision-making process and are indifferent to the nature of its embodiment. Most empirical studies of decision making, however, have focused

on the behavior of individuals. Relatively little attention has been given to the question of how decisions are, or should be, made by n-person groups. There are, of course, large literatures dealing with related topics such as the effects of group organization and communication channels on problem solving, and the effects of group pressures on individual behavior.

One generalization that it seems safe to make is that the decision-making performance of groups may be influenced by a number of factors that are not obviously related to decision quality in any straightforward way. Especially is this true when group members are required to resolve problems about which there exist conflicting views. As Helmer (1967) puts it:

> "Round-table discussions for such purposes have certain psychological drawbacks in that the outcome is apt to be a compromise between divergent views, arrived at all too often under the undue influence of certain factors inherent in the face-to-face situation. These may include such things as the purely specious persuasion of others by the member with the greatest supposed authority or even merely the loudest voice, an unwillingness to abandon publicly expressed opinions, and the bandwagon effect of majority opinion" (p. 9).

As one means of remedying these types of problems, and of providing a rationale by which to combine "expert" opinions, the Delphi method was created (Brown, 1968; Dalkey & Helmer, 1963; Helmer, 1967; Rescher, 1969). This technique requires each member of the group to write down his independent assessment of the problem or solution under study. The set of assessments is then revealed to all members but without identification of which particular assessment was made by which member. The pros and cons of each response are then openly debated and each member files a second assessment. Following n repetitions of this procedure, the median assessment is then adopted.

The Delphi procedure is reputed to be usable:

> "1) To determine what the operative values of a group are, what relative weight they have, what sorts of possible trade-offs obtain among them, and the like.
>
> 2) To explore the sphere of value criteriology, clarifying by what criteria the values of a group come to be brought to bear upon actual cases.
>
> 3) To discover divergences of value posture within a group and the existence of subgroups with aberrant value structures.

4) To serve as a tool for seeking out areas of value consensus--or agreement as to actions and preferences--that may exist even when there are conflicts of value.

5) To provide a tool for the third-party evaluation of conflicts of interest.

6) To assess the correctness of value ascriptions to given groups.

7) To assess the correctness of value judgments in the area of means-values" (Rescher, 1969, p. 17).

The use of a modified version of the Delphi technique is illustrated in a recent effort by O'Connor (1972) to apply expert judgment to the scaling of water quality. The problem was to assess the quality of water to be used (1) as a public supply, and (2) for the maintenance of a fish and wildlife population. Eight experts made iterative judgments as to the parameters to be included, the relative importance weights to be assigned, and the rules for combination of indices. Good consensus was obtained with respect to sets of judgment parameters and combination rules, but there was considerable disagreement on weightings. O'Connor found, however, that this disagreement was not critical in the development of the final indices.

An important feature of the Delphi technique is the fact that it provides a means for achieving group consensus without the need for the face-to-face discussion of issues which typifies most group problem-solving methods. This characteristic was exploited in the O'Connor study, where the experts were geographically widely separated and were never in direct communication with each other.

13.4 Computer-Based Decision Aids

The potential advantages to be gained from applying the general computational capabilities of digital computers to decision problems have been recognized for some time. Several writers have made very convincing arguments to the effect that both men and computers have something to offer to the decision-making process, and that the need is for the development of decision systems that assure a symbiotic coupling of the capabilities of man and machine (Briggs & Schum, 1965; Edwards, 1965b; Licklider, 1961; Shuford, 1965; Yntema & Klem, 1965; Yntema & Torgerson, 1961).

It is not difficult to imagine a computer system being used to aid a decision maker in the performance of essentially all of the aspects of decision making that we have considered in foregoing sections of this report. Such a system might provide the decision maker with a data base of facts or observations that are

relevant to his decision problem. It could serve as an extension of his own memory by keeping a record of factors that he had indicated he ought to "keep in mind" in making a decision. It could help him generate hypotheses, and to structure and present the decision space. It could help him discover what his preferences are and to express them in a quantitative way. It could provide graphical representations of the decision situation. It might (assuming a valid model of the decision problem) project the probable consequences of various action selections. It might serve as an interface between two or more decision makers' collaborating on the same problem and facilitate the application of group decision techniques. It could do whatever computation was required. It could prod the decision maker to consider aspects of the problem that he otherwise might overlook. It could suggest approaches or strategies that have been found to be useful in similar problem situations. It could make explicit to the decision maker (either by inference or by questioning of the decision maker himself) some aspects of the situation or the decision maker's thinking that otherwise would only be implicit. And so on.

It is in fact so easy to imagine ways in which the computer could be used as an aid for decision making that one can be seduced to thinking that the implementation of such capabilities is a straightforward thing. In some instances this is perhaps the case; in others, it assuredly is not. The important point is, however, that computer-based decision aids are being developed and quite sophisticated ones are likely to be operational in the near future. No training program for decision makers can afford to ignore this fact.

In a preceding section of this report some comments were made concerning simulation as an approach to training. Given the availability of computer systems to decision makers, another way that simulation may be used to advantage is as an operational decision aid. In this case the effects, or probable effects, of selecting specific action alternatives can be explored by the decision maker before he actually makes his choice (Ferguson & Jones, 1969). The projections or predictions of the aid will only be as good, of course, as is the model of the situation that produces them, and it is not necessarily the case that the use of such predictive aids will invariably lead to improved performance (Sidorsky & Mara, 1968). The potential for this type of simulation is great, however, and deserves more attention that it has received to date. At the very least, such an aid can be used to help determine what is possible and what is not, giving an accurate representation of the current state of affairs. The point is illustrated by an experimental decision aid designed to monitor and control maritime traffic (Elmalph, Prywes, & Gustafemo, 1967). The system was composed of a formatted data base, a set of "worker programs" which operated on the data base, and a query language which allowed the

user to interact with the data base on line. Information that could be extracted from the data base on request included" "(1) past, present, or future locations of ships, (2) the number, type, or names of ships in any geographic area of the North Atlantic at a past, present, or future time, and (3) how far is a ship from some particular place and if ordered to change course when can it get there?" (p. 206). The system could provide information on sets of ships satisfying some class description; for example, it could provide the distances of all ships of a given type, from a given destination, and the time required to reach that destination, assuming the necessary change in course. The system illustrates a nice allocation of function between man and machine. The computer does the bookkeeping and arithmetic, the man exercises judgment and makes choices. Hopefully, the choices that the man makes will be the better because of the bookkeeping and arithmetic that the machine does.

Two of the more prominent problem areas for which computer-based decision aids have been developed or planned are medicine and military tactics.

13.4.1 Computer-Based Aids for Medical Decision Making

Among the first investigators to attempt to apply modern decision theory to medical decision making were Ledley and Lusted (1959). During the subsequent fifteen years, many such applications of decision theoretic techniques were proposed and tried; and within the past ten years, several experimental computer-based systems have been developed for the purpose of facilitating various aspects of decision making in the medical context. Applications that have been explored include initial patient interviewing and symptom identification (Griest, Klein, & VanCura, 1973; Whitehead & Castleman, 1974), analysis organization and presentation of the results of laboratory tests (Button & Gambino, 1973), personality analysis (Kleinmuntz, 1968; Lusted, 1965), storage and retrieval of individual-patient data (Collen, 1970; Greene, 1969), on-demand provision to practitioners of clinical information (Siegel & Strom, 1972), automated and computer-aided diagnosis of medical problems (Cumberbatch & Heaps, 1973; Fisher, Fox, & Newman, 1973; Fleiss, Spitler, Cohen, & Endicott, 1972; Gledhill, Mathews & Mackay, 1972; Horrocks & deDombal, 1973; Jacquez, 1972; Loewick, 1965; Lusted, 1965; McGirr, 1969; Yeh, Betyar, & Hon, 1972), management and graphical representations of data to aid research in pharmacology and medicinal chemistry (Castleman, Russell, Webb, Hollister, Siegel, Zdonik, & Fram, 1974), modelling of physiological systems and exploration via simulation of the effects of alternative courses of treatment (Seigel & Farrell, 1973), and training (Feurzeig, 1964; Feurzeig, Munter, Swets, & Breen, 1964).

The results of one recent study of computer-assisted diagnosis are particularly relevant to the question of when expert judgment should or should not be used in the decision process. Leaper (1972, 1975) compared two methods of computer-assisted diagnosis of disorders for which abdominal pain was a primary symptom (e.g., appendicitis, diverticulitis, perforated ulcer). Computer-aided Bayesian diagnoses were performed using estimates of probabilities that were either (a) inferred from frequency data collected from 600 patients or (b) produced by a group of clinicians. The diagnoses that resulted from the computer-aided method that used the clinicians' probability estimates were marginally more accurate than those produced by unaided clinicians (82% versus 80%). The method that made use of probabilities inferred from incidence data, however, gave significantly more accurate results (91%). A secondary result of this study that is of some interest is the fact that most clinicians insisted on retaining their own probability estimates, even when those estimates were greatly different from the survey data and they had been informed of this fact.

These results strongly suggest that relative frequency data should be used as a basis for probability estimates in preference to expert opinions, if such data are available. The principle should not be applied, of course, without due regard for such factors as the size and representativeness of the samples from which the relative frequency data are obtained. As a general rule, the most defensible strategy in estimating probabilities would seem to be: use expert judgments only if a more objective method is not feasible, as would be the case when estimating the probabilities of very low-frequency events or events that are not reasonably thought of as "frequentistic" in nature.

13.4.2 Computer-Based Aids for Tactical Decision Making

Much has been written about the use of computer-based aids to facilitate decision making in the context of tactical operations (Alden, Levit, & Henke, 1973; Baker, 1970; Bennett, Degan, & Spiegel, 1964; Bowen, Feehrer, Nickerson, & Triggs, 1975; Bowen, Feehrer, Nickerson, Spooner, & Triggs, 1971; Bowen, Halpin, Long, Lukas, Mullarkey, & Triggs, 1973; Freedy, Weisbrod, May, Schwartz, & Wettman, 1973; Gagliardi, Hussey, Kaplan, & Matten, 1965; Hanes & Gebhard, 1966; Levit, Alden, & Henke, 1973; Levit, Alden, Erickson, & Heaton, 1974; Sidorsky & Simoneau, 1970). The extent to which such systems and aids have led to improved decision making is probably impossible to determine. It is easy to be critical of this work, however, because progress has certainly not been spectacular. And it may be that some of the decision-aiding efforts have been poorly conceived. But tactical decision making is complicated and not thoroughly understood. It is not surprising that there would be some false starts before significant progress is made on this problem. Even false starts can provide useful insights into a

problem, however; if nothing more, they should help to clarify what the dimensions of the problem are and to provide some clues concerning the requirements for a solution.

Bowen, Nickerson, Spooner, and Triggs (1970) have described several computer-based systems that have been, or are being, developed by the military services to aid the decision-making process in tactical situations. Among the systems that were reviewed are: the Army's Tactical Operations System (TOS)--in particular, TOS-7th Army--and Tactical Fire Direction System (TACFIRE), the Air Force and Marine Corps' Tactical Information Processing and Interpretation System (TIPI), the Air Force's Intelligence Data Handling System (IDHS), and the Navy's Integrated Operational Intelligence System (IOIS). These systems are intended to improve tactical decision making by facilitating data management and manipulation, message routing, display generation, report preparation, fire control, planning, resource allocation, and other tasks and functions that fall within the purview of tactical operations.

There are two motivations for bringing such systems into the tactical situation. One is to unburden the decision maker of tasks that are just as well performed by machines, and thereby make it possible for him to devote more time to those aspects of decision making that require human judgment and expertise. The other is to upgrade the quality and adequacy of the information on which decisions are based. This involves not only the problem of processing and integrating large amounts of information, but also that of packaging and presenting information in ways that are well-suited to the information-processing capabilities of the human being who must make use of it. How effectively existing or contemplated systems realize these objectives is difficult to determine with much precision.

It is not the purpose of this review to describe particular systems in detail. We will, however, consider briefly two systems as illustrative of those that have been developed, one intended for operational use, and one for use as a training instrument.

13.4.2.1 AESOP

An intensive program to develop an on-line information-control system of value to military decision makers in the planning of tactical and strategic resource allocations was begun by the Mitre Corporation in 1964. On completion in 1969, the prototype, called "An Evolutionary System for On-line Planning (AESOP) to emphasize its incremental approach to the generation of computer-based management and planning assistance, made available to system users a range of techniques which could aid in such diverse activities as data acquisition, aggregation, plan assessment and report preparation.

The AESOP system consists of two major parts. One of these is a set of capabilities for storing, modifying, retrieving, and displaying data, and for performing various sorts of symbolic and arithmetic manipulations with the aid of a flexible display-oriented user language, a light pen, typewriter and push-buttons. Details of these aspects are covered in a variety of program publications, the most informative of which are Bennett, Haines, and Summers (1965) and Summers and Bennett (1967).

The second part of the system consists of a set of simulated strategic and tactical military applications which provide a context for exercising the capabilities mentioned above. One of the more significant of these is that of a Tactical Air Control Center (TACC) in which the resource allocation tasks of a Fighter Section/Current Plans Division are simulated. Since this particular application also served as a testbed for the formal test and evaluation of AESOP principles, it provides the most comprehensive picture of the strengths and weaknesses of the system. The remainder of our current summary will relate to this application and to the results of evaluation studies. More detailed treatments of the simulation and evaluation can be found in Doughty (1967), Doughty and Feehrer (1969), and Doughty, Feehrer, Bachand and Green (1969).

As simulated in the AESOP program, the basic task of a Fighter Section revolves about the allocation (on request by higher headquarters) of tactical aircraft to each of three mission categories: (1) on-call close air support, (2) preplanned close air support, and (3) preplanned counter-air and interdiction. Under "normal" circumstances the total number of ready aircraft in near proximity to prescribed target areas is less than the number of aircraft requested, so the planner is forced to make tradeoffs relating to such factors as sortie rate, flying time, time over target, and probable degree of target destruction. The cumulative consequences of these tradeoffs are: (1) that some requests for support fail to be satisfied at all, (2) some requests fail to be satisfied on a timely basis, and (3) some requests, though satisfied on a timely basis, are not satisfied at the required level.

In this context, the tactical version of AESOP has two interrelated goals: (1) the elimination of much of the labor and inaccuracy associated with manual computation and display of ready resources, sortie rates, flying times and weapons' effects, and with the preparation of formal orders (Fragmentary Orders) to squadrons implicated in a planned allocation, and (2) facilitation of the problem-solving activity of decision makers, that is, of the judicious selection of squadrons, aircraft types, weapons categories, and so on.

For purposes of evaluation, the actual resource allocations produced by planners using the AESOP system were compared with

those produced by planners using a simulated version of the standard system in an integrated series of tactical exercises depicting the military maneuvers of loyalists and insurgents during a ten-day limited war. Experimental sessions began with briefings relating to orders of battle, political and military activity, and Joint Task Force requests for support of loyalist objectives. AESOP and Manual Planning teams then adjourned to commence allocation activities in response to the simulated JTF requests. The experiment ended each day with the (automated or manual) production of squadron Fragmentary Orders.

Each planner, whether operating the manual or AESOP system, was required to generate an allocation which represented, in his judgment, the best tradeoff among four criteria (listed in decreasing order of importance):

1. Satisfaction of requested level of damage
2. Satisfaction of requested time over target
3. Minimization of use of recycled aircraft (i.e., of sortie rate)
4. Minimization of (total) flying time.

The results of the evaluation study contained few surprises. In those aspects of planning activity for which AESOP provided direct assistance, performance of those using the system was superior. In those aspects for which assistance was not provided, planners in the two systems performed at approximately equal levels. The net performance of AESOP planners was superior to that of manual planners with respect to plan quality and production efficiency, a finding that must be assessed in light of the fact that the larger portion of the task was fairly routine and required little creative ability.

It is important to note that the AESOP system provided no formal procedural aids to the decision maker such as decision algorithms, linear programming solutions, etc. What benefits accrued to users of the system during the more creative phases of their task seemed to result from a combination of indirect factors. It appeared to be the case, for example, that planners could more easily comprehend the extent to which resources would be "strained" and, thereby, develop a better "feel" for the nominal form of their plan prior to its production. This appreciation for the difficulty of the problem with which they were faced on a particular day was materially aided by the concise nature of the displays provided by the system. Planners who used the system were in a much better position to monitor their own progress while solving the problem than were those whose appreciation of the demands of the situation had to be assembled from groups of formal documents.

It appeared also to be the case that, since the system performed routine aspects automatically, more time was available for creative activities and for reiteration of plans. On several occasions AESOP planners attempted successfully to produce series of allocations of progressively greater merit and stopped only when they were totally satisfied with their efforts.*

13.4.2.2 TACTRAIN

The Tactical Training (TACTRAIN) facility was developed by the Electric Boat Division of General Dynamics, partly as a demonstration that a modest computer with a CRT display could be employed in the training of decision-making skills and partly as an experimental tool for evaluation of alternative tactical display/interrogation formats. Details regarding computer and display equipment, software and tactical problem parameters used in the system and a specific configuration employed for system evaluation are discussed by Sidorsky and Simoneau (1970). The summary presentation below draws heavily on their discussion.

The TACTRAIN system provides an opportunity for the decision maker to take on the role of a commanding officer of a submarine on an ASW search-and-destroy mission. His specific task is to maneuver in such a way that he simultaneously maximizes the probability of destroying a simulated enemy ship and minimizes the probability that the enemy ship will destroy him. He chooses a maneuver by selecting a speed, a depth, a firing range, and a quantity of torpedoes, each from among five alternatives. The choices are constrained to be consistent with the operating characteristics of own and enemy ships, the parameters of own ship's weapons complement, and specific sound channel, topographic and bathythermal conditions. The maneuver implied by the alternatives that are chosen is then evaluated with respect to each of four criteria: (1) the probability that own ship would be able to detect the enemy ship, (2) the probability that the enemy ship would be able to detect own ship, (3) the probability that own ship would be able to destroy the enemy ship, given the maneuver and weapon characteristics, and (4) the probability that the enemy ship would be able to destroy own ship.

While solving a particular tactical problem, the officer can retrieve information stored in the system by interrogating the display with a light pen. Appropriate interrogations lead

---

*A quasi-linear program to aid this strategy at a formal level was later developed by Feehrer (1968) for the AESOP TACC planning activity.

to one of two categories of display: (1) prior to a "command decision," graphic displays of the "tactical effectiveness" associated with the choice of a particular alternative on each tactical dimension (speed, range, etc.) with respect to each of the four criteria--available prior to a command decision, and (2) alphanumeric displays revealing the outcome of the maneuver, the number of (quality) points to be assigned to the outcome, and the cumulative number of points acquired as of the end of the experimental trial in question--available following a command decision.

The developers of TACTRAIN see it making at least two valuable inputs to the learning process of the decision maker. First, it provides immediate knowledge of the consequences of a decision. The decision maker discovers very quickly whether he destroyed the enemy ship and whether his own ship was destroyed in the process. Moreover, he is provided with an arithmetic measure, however arbitrarily derived, of his cumulative performance.

Second, the decision maker is provided, via the display, with a graphic portrayal of the interactions between tactical and environmental variables and their relationship to tactical effectiveness as represented by detection/counter detection and hit/miss outcomes. And, inasmuch as the tactical problem unfolds over time, the decision maker also gains an appreciation for the changing complexities of these interactions and for the need for timeliness in his decision.

13.4.3 Computer-Based Decision Aids and Training

It seems highly probable that many attempts to develop computer-based decision aids will fail in the sense that the aids that are produced will not measure up to the expectations of their developers. This is not necessarily failure in a larger view, however, if these attempts lead to a better understanding of the decision-making process--as one might reasonably hope that they will. To the extent that these efforts do lead to new insights into various aspects of the decision-making process, they will have direct impact on training curricula.

To the extent that specific systems prove to be effective aids in operational situations, they will constitute new tools with which decision makers will have to work. Thus, their existence will represent a new training need, namely the need to train the users of these aids.

Perhaps the most challenging way in which the development of increasingly sophisticated computer-based systems relates to training is in the potential that these systems represent for providing training for their users. Critics of the idea of

computer-assisted instruction can correctly point out that the results of endeavors in this area have not measured up to the expectations that were fostered by many of the early enthusiasts for this use of computers. Very real progress in the area is being made, however, and it may prove to be the case that the early enthusiasts erred only in failing to appreciate the difficulty of some of the problems that had to be solved and the time that would be required to solve them. There is no question but that computer systems that are intended to be used by people interactively on complex problem-solving tasks can be given the capability to provide much of the training that is required, both to initiate users and to bring users from neophyte to expert status. The potential gains to be realized by building such training capabilities into operational systems suggest that this possibility is worth far more attention than it has yet received.

ACKNOWLEDGMENTS

We wish to acknowledge, and express our appreciation for, the assistance of several people in this project. Charlene Long, Walter Hawkins and George Lukas helped to locate and screen articles. Ann Rollins wrote the computer programs that were used to generate the results reported in Section 8.6. Michael Samet made several useful comments on a draft of the report. Florence Maurer typed the manuscript and helped over the course of the project in many ways. Sheila Allen put the references on a computer file and thereby facilitated the preparation of the bibliography.

Abt, C.C. Serious games. New York: The Viking Press, 1970.

Ackoff, R.L. Management misinformation systems. Management Science, 1967, 14, b-147-156.

Aczel, J., & Pfanzagl, J. Remarks on the measurement of subjective probability and information. Metrika, 1966, 11, 81-105.

Adelson, M. Human decisions in command control centers. Annals New York Academy of Science, 1961, 89, 726-731.

Alberoni, F. Contribution to the study of subjective probability. Part I. Journal of General Psychology, 1962, 66, 241-264.

Alden, D., Levit, R. Henke, A. Development of operational decision aids for naval task force application. Honeywell TR 7201-3048, 1973.

Alexander, L.T. A demonstration of the use of simulation in the training of school administrators. New York: City University of New York, Division of Teacher Education, 1967, 114, AD 014157.

Algea, C.W. A development of a conceptual framework of the driving task. Human Factors, 1964, 6, 375-385.

Anderson, N. H. A simple model for information integration. In R. P. Abelson, et al (Eds.), Theories of cognitive constancy: A source book. Chicago: Rand McNally, 1968.

Anderson, N. H. Comment on an analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. Psychology Bulletin, 1959, 72, 63-65.

Andrews, R. S., Jr. & Ringel, S. Certitude judgments and accuracy of information assimilation from visual displays. U.S. Army Personnel Research Office Technical Research Note No. 145, 1964

Armer, P. The use of computers. In E. S. Quade (Ed.), Analysis for military decisions. Chicago: Rand-McNally, 1964.

Attneave, F. Psychological probability as a function of experienced frequency. Journal of Experimental Psychology, 1953, 46, 81-86.

Bacon, F. The new organon. In H.G. Dick & Francis Bacon (Eds.), Selected Writings, Modern Library Edition, New York: Random House, 1955.

Baker, F.P. The internal organization of computer models of cognitive behavior. Behavioral Science, 1967, 12, 156-161.

Baker, J.D. A technique for obtaining non-dichotomous measures of short term memory. Decision Sciences Laboratory Report, ESD-TR-64-674, 1968.

Baker, J.D. The uncertain student and the understanding computer. Paper presented at a colloquium on Programmed Learning Research, Nice, France, 1968.

Baker, J. D. Quantitative modelling of performance in information systems. Ergonomics, 1970, 13, 645-664.

Baker, J.D., McKendry, J.M., & Mace, D.J. Certitude judgments in an operational environment. U.S. Army Behavioral Science Research Laboratory, Technical Research Note 200, November, 1968.

Bartlett, F.C. Thinking, an experimental and social study. London: Allen and Unwin, 1958.

Beach, L.R., & Swensson, R.G. Intuitive estimation of means. Psychonomic Science, 1966, 5, 161-162.

Becker, G.M. Sequential decision making: Wald's model and estimates of parameters. Journal of Experimental Psychology, 1958, 55, 628-636.

Becker, G.M., Degroot, M.H., & Marscak, J. Stochastic models of choice behavior. Behavioral Science, 1963, 8, 41-55.

Becker, G.M., & McClintock, G.G. Value : Behavioral decision theory. Annual Review Psychology, 1967, 18, 239-286.

Bellman, R.E., Clark, C.E., Malcolm, D.G., Graft, D.J., & Ricciardi, F.M. On the construction of a multistage, multiperson business game. Operations Research, 1957, 5, 469-503.

Benington, H.D., Military information - recently and presently. In E. Bennett, J. Degan, and J. Spiegel, (Eds.), Military information systems. Praeger, 1964.

Bennett, E., Degan, J., & Spiegel, J. Military information systems: The design of computer-aided systems for command. New York: Praeger, 1964.

Bennett, E. M., Haines, E. C., & Summers, J. K. AESOP: A prototype for on-line user control. AFIPS Conference Proceedings: Fall Joint Computer Conference, 1965, 27, 435-455.

Bentham, J. An introduction to the principles of morals and legislation, (originally published in 1789). In E.A. Burtt (Ed.), The english philosophers from Bacon to Mill, Modern Library Edition, New York: Random House, 1939.

Bernoulli, D., Specimen theoriae norae de mensura sortis. Comentarii Acade,oae Scientiarum Imperiales Petropolitanae, 1738, 5, 175-192, (Translated by L. Sommer in Econometrica, 1954, 22, 23-26).

Beville, J., Wagner, J. H., & Zannetos, Z. S. The development of an interactive graphical risk analysis system. M. I. T., Sloan School of Management, 1970, 502-570.

Birdsall, T. G. & Roberts, R. A. Theory of signal detectability: Deferred-decision theory. The Journal of the Acoustical Society of America, 1965, 37, 1064-1074.

Blackwell, D., & Girshick, M. Theory of games and statistical decisions. New York: Wiley, 1954.

Blaiwes, A. S. Some training factors related to procedural performance. Journal of Applied Psychology, 1973, 58, 214-218.

Bond, N. A., Jr. & Rigney, J. Bayesian aspects of trouble shooting behavior. Human Factors, 1966, 8, 377-385.

Boole, G. An investigation of the laws of thought. Macmillan, 1854.

Bowen, R. J., Feehrer, C. E., Nickerson, R. S., Spooner, R. L., & Triggs, T. J. Army tactical intelligence: An examination of computer-based displays and a structured decision procedure as aids to the intelligence analyst. BBN Report 2073 under Contract DAHC-19-69-C-0020, US Army Behavior and Systems Research Lab, 1971.

Bowen, R. J., Feehrer, C. E., Nickerson, R. S., & Triggs, T. J. Computer-based displays as aids in the production of army tactical intelligence. USA Research Institute for the Behavioral and Social Sciences Technical Paper 258, February 1975.

Bowen, R. J., Halpin, J. A., Long, C. A., Lukas, G., Mullarkey, M. M., & Triggs, T. J. Decision flow diagrams and data aggregation in army tactical intelligence. BBN Report # 2570 to USA Research Institute for the Behavioral and Social Sciences, June 1973.

Bowen, R. J., Nickerson, R. S., Spooner, R. L., and Triggs, T. J. Army tactical intelligence: An investigation of current methods of production and possible computer-based

innovations. BBN Report No. 1945, 16 Sept. 1970, to USA Behavior and Systems Research Laboratory under Contract DAHC-19-69-c-0020.

Brier, G.W. Verification of forecasts expressed in terms of probability. Monthly Weather Review, 1950, 78, 1-3.

Briggs, G. E., & Schum, D. A. Automated Bayesian hypothesis-selection in a simulated threat-diagnosis system. In J. Spiegel & D. E. Walker (Eds.), Information systems sciences: Proceedings of the second congress. Washington, D.C.: Spartan Books, 1965, 169-176.

Brody, N. The effect of commitment to correct and incorrect decisions on confidence in a sequential decision task. American Journal of Psychology, 1965, 78, 251-256.

Bruner, J. S. On perceptual readiness. Psychological Review, 1957, 64, 123-152.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: Wiley, 1956.

Brown, B. Delphi process. The Rand Corporation, p-3925, 1968.

Brunswick, E. The conceptual framework of psychology. Chicago: University of Chicago Press, 1952.

Brunswick, E. Perception and the representative design of experiments. Berkeley: University of California Press, 1956.

Button, K. F., & Gambino, S. R. Laboratory diagnosis by computer. Computers in Biology and Medicine, 1973, 3, 131-136.

Carr, J. W., III, Prywes, N. S., Bursky, P., Linzen, S., & Lull, R. Man-computer problem solving in real-time naval duels. University of Pennsylvania, Moore School Report No.: 71-08, 1970.

Castleman, P. A., Russell, C. H., Webb, F. N., Hollister, C. A., Siegel, J. R., Zdonik, S. R., & Fram, D. M. The implementation of the PROPHET system. Proceedings of National Computer Conference, 1974.

Chenzoff, A., Crittenden, R. L., Flores, I., & Tolcott, M. A. Human decision making as related to air surveillance systems: Technical Report No. 1, A survey of literature and current research. AFCCDD-TR-60-25, June 10, 1960. Issued by Dunlap and Associates.

Clarkson, G. P. E. A model of the trust investment process. In E. A. Feigenbaum & J. Feldman (Eds.), Computers and

thought. New York: McGraw-Hill, 1963.

Cohen, K. J., & Rhenman, E. The role of management games in education and research. Management Science, 1961, 7, 131-166.

Cohen, J., Chesnick, E. I., & Haran, D. Confirmation of inertial - PSI effect in sequential choice and decision. British Journal of Psychology, 1972, 63, 41-46.

Cohen, J, Dearnaley, E. J., & Hansel, C. E. M. Risk and hazard. Oper. Research Quarterly, 1956, 7, 3, 67-82.

Cohen, J., Dearnaley, E. J., & Hansel, C. E. M. Skill with chance: Variations in estimates of skill with an increasing element of chance. British Journal of Psychology, 1958, 49, 319-323.

Cohen, J., & Hansen, C. E. M. Risk and gambling: The study of subjective probability. New York: Philosophical Library, 1956.

Cohen, M. R., A preface to logic. Holt, 1944.

Collen, M. F. (Ed.) Medical Information Systems, Springfield, Va., National Technical Information Services, 1970.

Coombs, C.H., Inconsistency of preferences: A test of unfolding theory. In W. Edwards & A. Tversky (Eds.), Decision making. Baltimore: Penguin, 1967.

Coombs, C. H., Bezembinder, T. G., Goode, F. M. Testing expectation theories of decision making without utility or subjective probability. Journal of Mathematical Psychology, 1967, 4, 72-103.

Coombs, C. H., Dawes, R. M., & Tversky, A. Mathematical Psychology: An Elementary Introduction, New Jersey: Prentice-Hall, 1970

Cloot, P. L., Management information systems- Can computers help? The Computer Bulletin, 1968, 11, 276-281.

Cumberbatch, J. & Heaps, H. S. Applications of a non-Bayesian approach to computer-aided diagnosis for upper abdominal pain. International Journal of Biomedical Computing, 1973, 4, 105-115.

Dale, H. C. A. Weighing evidence: An attempt to assess the efficiency of the human operator. Ergonomics, 1968, 11, 215-230.

Dalkey, N. & Helmer, O. An experimental application of the Delphi method to the use of experts. Management Science, 1963, 9, 458-467.

Damas, P. A., Goodman, B. C., & Peterson, C. R. Bayes theorem: Response scales and feedback. University of Michigan project number NR 197-014, Sept. 1972.

Davidson, D., Suppes, P., & Siegel, S. Decision making: An experimental Approach. Stanford: Stanford University Press, 1957.

Davies, I. Get immediate relief with an algorithm. Psychology Today, April, 1970, 53-69.

Dawes, R. M., & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

DeFinetti, B. Does it make sense to speak of "good probability appraisers"? In I. J. Good, (Ed.) The scientist speculates: An anthology of partly-baked ideas. New York: Basic Books, 1962.

Dillman, D. H., & Cook, D. L. Simulation in the training of R&D project managers. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, Calif., Feb. 8, 1969, 20 p. ED 209373.

Dodson, J. D., Simulation system design for a TEAS simulation research facility. Report AFCRL 1112 and PRC 194. Planning Research Corporation, Los Angeles, November 15, 1961.

Doughty, J. M. The AESOP testbed: Test series I/2. In D. E. Walker (Ed.), Information system science and technology. Washington D.C.: Thompson Book Co., 69-86, 1967.

Doughty, J. M., & Feehrer, C. E. The AESOP testbed. Test Series I/2 (Summary Report). Mitre Technical Report No. 848, 1969.

Doughty, J. M., Feehrer, C. E., Bachand, T. E., & Green, E. R. The AESOP testbed, Test series I/2: Supplementary technical information. Vol. 1, System descriptions, test procedures, and data analysis Mitre Technical Report No. 824, 1969.

Dreyfus, S. Dynamic Programming. In R. L. Ackoff, (Ed.), Progress in operations research. New York: Wiley, 1961.

Drucker, P. F. The effective decision. Harvard Business Review, 1967, 45, 92-98.

Duncker, K. On problem-solving. In P.C. Wason & P.N. Johnson-Laird (Eds.), Thinking and Reasoning, Baltimore: Penguin Books, 1968, 28-43.

Dutton, J. M. & Starbuck, W. H. Finding Charlie's run time estimator. In J. M. Dutton & W. H. Starbuck (Eds.), Computer simulation of human behavior. New York: Wiley, 1971.

Edwards, W. The theory of decision making. Psychological Bulletin, 1954, 51, 380-417.

Edwards, W. The use of statistical decision functions in making practical decisions. Paper given at Symposium on Air Force Human Engineering, Personnel, and Training Research, November 14-16, 1955, Washington, D.c. Publication 455, NAS-NRC, 1956.

Edwards, W. Behavioral decision theory. Annual Review of Psychology, 1961, 12, 473-493.

Edwards, W. Probabilistic information processin by men, machines and man-machine systems. System Development Corporation TM-1418/000/01 August 1963. (Paper presented at 18th International Congress of Psychology, Washington, D.C., August 1963.

Edwards, W. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. Journal of Mathematical Psychology, 1965a, 2, 312-329.

Edwards, W. Probabilistic information processing system for diagnosis and action selection. In J. Spiegel & D. Walker (Eds.), Information system sciences, Proceedings of the Second Congress. Washington, D.C.: Spartan Books, Inc., 1965b, 141-155.

Edwards, W. Dynamic decision theory and probabilistic information processing. Human Factors, 1967, 4, 59-73.

Edwards, W. A bibliography of research on behavioral decision process to 1968. Memorandum Report No. 7, Human Performance Center, University of Michigan, Ann Arbor, Michigan, 1969.

Edwards, W. Research on the technology of inference and decision. University of Michigan NTIS AD 771 463, 1973.

Edwards, W., Lindman, H., and Phillips, L. D. Emerging technologies for making decisions. New Directions in Psychology II. New York: Holt, Rinehart, & Winston, 1965, 261-325. Edwards, W., Phillips, L. D., Hays, W. L., & Goodman, B. C. Probabilistic information processing systems: Design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 248-265.

Edwards, W., & Slovic, P. Seeking information to reduce the risk of decisions. American Journal of Psychology, 1965, 78, 188-197.

Edwards, W., & Tversky, A. Decision Making. Middlesex: Penguin Books, 1967.

Egan, J. P., Recognition memory and the operating characteristic. AFCRL-TN-58-51, Hearing and Communication Lab, Indiana University, Bloomington, Indiana, 1958.

Ellsberg, D. Risk, ambiguity and the Savage axioms. Rand Corp. memo no. P-2173, 1961.

Elmaleh, J. S., Prywes, N. S., and Gustafemo, J. F. Conversation with computers in the operation of sea traffic control. IEEE International Convention Record, Part 9, 1967, 203-209.

Erlick, D. E., Absolute judgments of discrete quantities randomly distributed over time. Journal of Experimental Psychology, 1964, 67, 475-482.

Estes, W. K. A descriptive approach to the dynamics of choice behavior. Behavioral Science, 1961, 6, 177-184.

Evans, J. R., & Cody, J. J. Transfer of decision-making skills learned in a counseling-like setting to similar and dissimilar situations. Journal of Counseling Psychology, 1969, 16, 427-432.

Feehrer, C. E. TACC Planning Aid. The Mitre Corp. Memo No. D74-M-473, February 6, 1968.

Feigenbaum, E. A. & Feldman, J. (Eds.). Computers and thought. New York: McGraw-Hill, Inc. 1963, 535 pages.

Ferguson, R. L. & Jones, C. H. A computer aided decision system. Management Science, 1969, 15, 10, B-550 - B-561.

Festinger, L. A theory of cognitive dissonance. Evanston: Row, Peterson, 1957.

Feurzeig, W. Conversational teaching machine. Datamation Magazine, June, 1964.

Feurzeig, W., Munter, P., Swets, J., & Breen, M. Computer-aided teaching in medical diagnosis. Journal of Medical Education, 1964, 39, 746-754.

Fischer, G. W. Four methods for assessing multi-attribute utilities: An experimental validation. University of Michigan project number NR 197-014, September 1972.

Fischer, G.W., & Peterson, C. R., Ratio versus magnitude estimates of importance factors. University of Michigan Report No. 037230-3-T, June, 1972.

Fishburn, P. Methods of estimating additive utilities. Management Science, 1967, 13, 435-453.

Fisher, M., Fox, R. I., & Newman, A. Computer diagnosis of the acutely ill patient with fever and rash. International Journal of Dermatology, 1973, 12, 59-63.

Fleiss, J. L., Spitzer, R. L., Cohen, J., & Endicott, J. Three computer diagnosis methods compared. Archives of General Psychiatry, 1972, 27, 643-649.

Fleming, R. A. The processing of conflicting information in a simulated tactical decision-making task. Human Factors, 1970, 12, 375-385.

Freedy, A., May, D., Weisbrod, R., & Weltman, G. Adaptive computer aiding in dynamic decision process: Scenario generation by elicited expert probabilities. Perceptronics Inc., Technical Report No. PTR-1016-74-5 (II), may 1, 1974.

Freedy, A., Weisbrod, R., May, D., Schwartz, S., & Weltman, G. Adaptive computer aiding in dynamic decision process. Technical Report PTR-73-101 Perceptronics, NTIS-AD 769 113, 1973.

Fried, L. S., & Peterson, C. R. Information seeking: Optional vs. fixed stopping. Journal of Experimental Psychology, 1969, 80, 525-529.

Funaro, J. F. An empirical analysis of five descriptive models for cascaded inference. Rice University, Department of Psychology Research Report Series, Report No. 74-2, May 1974.

Gagliardi, U. O., Hussey, R. A., Kaplan, T. T. & Matteis, R. J. Man-computer interactions in idealized tactical problem solving. Nonr-3602(00), 1965.

Geller, E. S. & Pitz, G. F. Confidence and decision speed in the revision of opinion. Organizational Behavior and Human Performance, 1968, 3, 190-201.

Gettys, C. F., Kelly, C. W., & Peterson, C. R. Best guess hypothesis in multi-stage inference. Organizational Behavior and Human Performance, 1973, 10, 364-373.

Gettys, C. F. & Wilke, T. A. The application of Bayes's theorem when the true data state is uncertain. Organizational Behavior and Human Performance, 1969, 4, 125-141.

Gibson, R. S. & Nicol, E. H. The modifiability of decisions made in a changing environment. ESD-TR-64-657, December, 1964.

Gledhill, V. X., Mathews, J. D., and Mackay, J. R. Computer-aided diagnosis: A study of bronchitis. Methods of Information in Medicine, 1972, II, 228-232.

Good, I. J. Rational decisions. J. Royal Statistical Society, Series B, 1952, 14, 107-114.

Gorry, G. A. The development of managerial models. M.I.T., Alfred P. Sloan School of Management, 1970.

Grabitz, H. J., & Jochen, H. An evaluation of confirming and disconfirming information in decision making (Grier). Archiv fur Psychologie, 1972, 124, 133-144.

Green, D. M., & Swets, J.A. Signal detection theory and psychophysics. New York: Robert E. Krieger Publishing Co., Inc., 1974. (Originally published, 1966.)

Green, P. E., Halbert, M. H. and Minas, J. S. An experiment in information buying. Journal of Advertising Research, 1964, 4, 17-23.

Greene, R. A. A computer-based system for medical record keeping by physicians. Harvard University, 1969.

Greist, J. H., Klein, M. H., & Van Cura, L. J. A computer interview for psychiatric patients target symptoms. Archives of General Psychiatry, 1973, 29, 247-253.

Guilford, J. P. Intellectual resources and their values as seen by scientists. In C. W. Taylor & F. Barron (Eds.), Scientific creativity: Its recognition and development. New York: Wiley, 1963.

Hammell, T.J., and Mara, T. D. Application of decision making and team training research to operational training: A translative technique. NAVTRADEVCEN 68-C-024201, 1970.

Hammond, J. S. III. Better decisions with preference theory. Harvard Business Review, 1967, 45, 123-141.

Hanes, R. M., & Gebhard, J. W. The computer's role in command decision. U.S. Naval Institute Proceedings, 1966, 92, 60-68.

Hayes, J. R. M. Human data processing limits in decision making. In E. Bennett (Ed.), Information system science and engineering. Proceedings of the First Congress on the Information Systems Sciences, New York: McGraw-Hill, 1964.

Haygood, R. C., & Bourne, L. E., Jr. Attribute and rule-learning aspects of conceptual behavior. Psychological Review, 1965, 72, 175-195.

Haythorn, W. W. Human factors in system research. Rand Corp. Report No. P-2337, June 1961. (To be a chapter in E. M. Bennett (Ed.), Human factors in technology.

Helmer, O. Systematic use of expert opinions. Rand Corp. memo No. P-3721, 1967.

Henke, A. H., Alden, D. G., & Levit, R. A. An analysis of human cognitive styles. Honeywell, Inc., Document F0141, 2 volumes, 1972.

Henle, M. On the relation between logic and thinking. Psychology Review, 1962, 69, 366-378.

Herman, L. M., Ornstein, G. N., & Bahrick, H. P. Operator decision performance using probablistic displays of object location. IEEE Transactions on Human Factors in Electronics, 1964, 5, 13-19.

Hill, J., & Martin, J. R. Training for educational decision making. Journal of Teacher Education, 1971, 22(4), 443-447.

Hoepfl, R. T., & Huber, G. P. A study of self-explicated utility models. Behavioral Science, 1970, 15, 408-414.

Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 57, 116.

Hoffman, P. J. Cue-consistency and configurality in human judgment. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley, 1968.

Hoffman, P.J. & Peterson, C. R. A scoring rule to train probability assessors. University of Michigan Project No. NR 197-014, 1972.

Hofstatter, P. R. Uber die Schatzung von gruppeneigenschaften. Zeitschrift fur Psychologie, 1939, 145, 1-44.

Hogarth, R. M. Process tracing in clinical judgment. Behavioral Science, 1974, 19, 248-313.

Horabin, I. Algorithms. Improving Human Performance: A Research Quarterly, 1972, 1, 28-41.

Horrocks, J. C., & de Dombal, F. T. Human and computer-aided diagnosis of 'dyspepsia', British Journal of Surgery, 1973, 60, 910.

Howard, R. A. The foundations of decision analysis. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 211-219.

Howard, R. A., Matteson, J. E., & North, D. W. The decision to seed hurricanes. Science, 1972, 176, 1191-1202.

Howell, W.C. Task characteristics in sequential decision behavior. Journal of Experimental Psychology, 1966, 71, 124-131.

Howell, W. C., & Gettys, C. F. Some principles for design of decision systems: A review of the final phase of research on a command control system simulation. USAF AMRL Tech. Report No. 68-158, 1968.

Huber, G., Sahney, V., & Ford, D. A study of subjective evaluation models. Behavioral Science, 1969, 14, 463-489.

Irwin, F. W. & Smith, W. A. S. Value, cost, and information as determiners of decision. Journal of Experimental Psychology, 1957, 54, 229-232.

Jacques, J. A. (Ed.) Computer diagnosis and diagnostic methods. Springfield, Illinois, Charles C. Thomas, 1972.

Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. Journal of Experimental Psychology, 1951, 41, 291-297.

Jensen, F. A., & Peterson, C. R. Psychological effects of proper scoring rules, Organizational Behavior and Human Performance, 1973, 9, 307-317.

Johnson, E. M. Numerical encoding of qualitive expressions of uncertainty. U.S. Army Research Institute for the Behavioral and Social Sciences Technical Paper No. 250, December 1973.

Johnson, E. M. The effect of data souce reliability on intuitive inference. ARI Technical Paper No. 251, AD 784097, July 1974.

Johnson, E. M., Cavanagh, R. C., Spooner, R. L., & Samet, M. G. Utilization of reliability measurements in Bayesian inference: Models and human performance. IEEE Transactions on Reliability, 1973, r-22, 176-183.

Kaplan, R. J., & Newman, J. R. A study in probabilistic information processing (PIP). System Development Corporation, TM-1150/000/00, April 1963.

Kaplan, R. J., & Newman, J. R. Studies in probabilistic information processing. IEEE Transactions in Human Factors in Electronics, 1966, 7, 49-63.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness, Cognitive Psychology, 1972, 3, 430-454.

Kahneman, D. & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Kanarick, A. F. The learning, retention and transfer of decision making. Paper presented at meetings on Learning, Retention and Transfer, sponsored by Honeywell Systems and Research Center and the Naval Training Devices Center, Winter Park, Florida, February 1969.

Kanarick, A. F., Huntington, J. M., & Petersen, R. C. Multi-source information acquisition with optional stopping. Human Factors, 1969, 11, 379-386.

Kant, I. Introduction to logic and essay on the mistaken subtilty of the four figures. (T. K. Abbott, trans.), Longman's, Green, 1885.

Kepner, C. H., & Tregoe, B. B. The rational manager: A systematic approach to problem solving and decision making. New York: McGraw Hill, 1965.

Kibbee, J. M. Dress rehearsel for decision making: The growing use of business games. The Management Review, 1959, 48, 4-73.

Kleinmutz, B. The processing of clinical information by men and machines. In Kleinmutz, B. (Ed.), Formal representation of human judgment. New York: Wiley, 1968.

Kneppreth, N. P., Gustafson, D. H., Leifer, R. P., & Johnson, E. M. Techniques for assessment of worth. U.S. Army Research Institute for the Behavioral and Social Sciences Technical Paper No. 254, 1974.

Krantz, D.H., & Tversky, A. Conjoint - measurement analysis of composition rules in psychology. Psychology Review, 1971, 8, 157-169.

Lathrop, R. G. Perceived variability. Journal of Experimental Psychology. 1967, 73, 498-502.

Leaper, D. J. Computer-assisted diagnosis of abdominal pain using "estimates" of data provided by clinicians and survey data obtained from real life. In T. J. Triggs & R. M. Pickett (Eds.), Human factors in health care, Lexington,

Massachusetts: D.C. Heath and Company, 1975.

Leaper, D. J., de Dombal, F. T., Horrocks, J. C., & Otaniland, J. R. Computer-assisted diagnosis of abdominal pain using `estimates´ provided by clinicians. British Journal of Surgery, 1972, 59, 897-898.

Ledley, R. S., & Lusted, L. B. Reasoning foundations of medical diagnosis. Science, 1959, 130, 9-21.

Lee, W. Decision theory and human behavior. New York: John Wiley & Sons, 1971.

Levine, J. M., & Samet, M. G. Information seeking with multiple sources of conflicting and unreliable information. Human Factors, 1973, 15, 407-419.

Levine, J. M., Samet, M. G., & Brahlek, R. E. Information seeking with input pacing and multiple decision opportunities. Human Factors, 1974, 16, 384-394.

Levit, R. A., Alden, D. G., Erickson, J. M., & Heaton, B. J. Development and application of a decision aid for tactical control of battlefield operations. Vol. 1: A conceptual structure for decision support in tactical operations systems. Honeywell, Inc. Contract number DAHC 19-73-C-0069, 1974.

Levit, R. A., Alden, D. G., & Henke, A. H. Development and application of a decision aid for tactical control of battlefield operations. Honeywell Document No. F6049-AA, 1973.

Lichtenstein, S. & Newman, J.R. Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 1967, 9, 563-564.

Lichtenstein, S. & Slovic, P. Reversals of preferences between bids and choices in gambling decisions. Journal of Experimental Psychology, 1971, 89, 46-55.

Lichtenstein, S., Slovic, P. & Zink, D. Effect of instruction in expected value on optimality of gambling decisions. Journal of Experimental Psychology, 1969, 79, 236-240.

Licklider, J. C. R. Man-computer symbiosis. IRE Transactions on Human Factors in Electronics, March 1960, 4-11.

Lindman, H. Inconsistent preferences among gambles. Indiana Mathematical Psychology Program Report 70-2, 1970.

Lodwick, G. S. A probabilistic approach to diagnosis of bone tumors. Radiologic Clinics of North America, 1965, 3,

487-497.

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

Luce, R. D. & Raiffa, H. Games and decisions: Introduction and critical survey. New York: Wiley, 1957.

Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

Lusted, L. B. Computer techniques in medical diagnosis. In R. W. Stacy and B. Waxman (Eds.), Computers in biomedical research, Volume 1, 1965, New York: Academic Press, 319-338.

MacCrimmon, K. R. Decision making among multiple-attribute alternatives: A survey and consolidated approach. Rand Corporation Memo RM 4823-ARPA, December 1968.

Mackworth, N. H. Originality. American Psychologist, 1965, 20, 51-66.

Maier, N. R. F. Reasoning in humans. II. The solution of a problem and its appearance in consciousness. Journal of Comparative Psychology, 1931, 12, 181-194.

Marschak, J. Rational behavior, uncertain prospects and measurable utility. Econometrica, 1950, 18, 111-141.

Martin, E. W., Jr. Teaching executives via simulation. Business Horizon, 1959, 2, 100-109.

McGirr, E. M. Computers in clinical diagnosis. In J. Rose (Ed.), Computers in medicine, 1969, London: J.A. Churchill, 19-29.

Meadow, C. L. & Ness, D. N. A case study of the use of a decision support system. M.I.T. Report No. 674-73, 1973.

Mill, J. S. A system of logic. Harper, 8th edition, 1874.

Miller, D. W. & Starr, M. K. The structure of human decisions. New Jersey: Prentice-Hall, Inc., 1967.

Miller, L.W., Kaplan, R.J., & Edwards, W. JUDGE: A value-judgment-based tactical command system. Organizational Behavior and Human Performance, 1967, 2, 329-374.

Minsky, M. Steps toward artificial intelligence. In E. A. Feigenbaum & J. Feldman (Eds.), Computers and thought. New York: McGraw-Hill, Inc., 1963, 406-450.

Morgan, J. J. B., & Morton, J. T. The distortion of syllogistic reasoning produced by personal convictions, Journal of Social Psychology, 1944, 20, 39-59.

Morin, R. E. Strategies in games with saddle-points. Psychological Report, 1960, 7, 479-485.

Morton, M. S. S. Decision support systems: the design process. Sloan School of Management, Report No. 686-73, 1973.

Neisser, U., & Weene, P. Hierarchies in concept attainment. Journal of Experimental Psychology, 1962, 64, 640-645.

Neyman, J., & Pearson, E. S. The testing of statistical hypotheses in relation to probability a priori. Proceedings of the Cambridge Philosophical Society, 1933, 29, 492-510.

Nickerson, R. S. & McGoldrick, C. C. Confidence, correctness, and difficulty with non-psychophysical comparative judgments. Perceptual and Motor Skills, 1963, 17, 159-167.

Nilsson, N.J. Problem-solving methods in artificial intelligence. McGraw-Hill, Inc., 1971.

North, D. W. A tutorial introduction to decision theory. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 200-210.

O'Connor, M. F. The assessment of worth for multi-attributed alternatives: A survey. Unpublished report, University of Michigan, 1972.

Organist, W. E. Use of subjective probabilities as a response technique in multiple choice behavior. Paper presented at the Fifth Annual Meeting of the Psychonomic Society, Ontario, Canada, October, 17, 1964.

Organist, W. E., & Shuford, E. H. Bayesian management systems. Paper presented at the University of Michigan, Second Bayesian Systems Conference, May, 1964.

Pascal, B. Thoughts. In C.W. Eliot (Ed.), The Harvard classics, 1910, 48, New York: Collier.

Paxson, E. W. War gaming. Rand Corp. memo no. RM-3489-PR, 1963.

Payne, J. W., Alternative approaches to decision making under risk: Moments versus risk dimensions, Psychological Bulletin, 1973, 80, 439-453.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Peterson, C. R., & DuCharme, W. M. A primacy effect in subjective probability revision. Journal of Experimental Psychology, 1967, 73, 61-65.

Peterson, C. R., et al. Handbook for decision analysis. Decisions and Designs, Inc., 1973, ARPA/ONR No. Nonr-N00014-73C-0149.

Peterson, C. R. & Miller,, A. J. Sensitivity of subjective probability revision. Journal of Experimental Psychology, 1965, 70, 117-121.

Peterson, C. R., & Phillips, L. D. Revision of continuous probability distributions. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 19-22.

Peterson, C. R. Schneider, R. J., & Miller, A. J. Sample size and the revision of subjective probability. Journal of Experimental Psychology, 1965, 69, 522-527.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. The theory of signal detectability, IRE Transactions, 1954, PGIT-Y, 171-212.

Phillips, L. D. & Edwards, W. Conservation in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Pitz, G.F. Sample size, likelihood and confidence in a decision. Psychonomic Science, 1967, 8 (6), 257-258.

Pitz, G. F. Information seeking when available information is limited. Journal of Experimental Psychology, 1968, 76, 25-34.

Pitz, G.F. An inertia effect (resistance to change) in the revision of opinion. Canadian Journal of Psychology, 1969, 23, 24-33.

Pitz, G. F., Downing, L., & Reinhold, H. Sequential effects in the revision of subjective probabilities. Canadian Journal of Psychology, 1967, 21, 381-393.

Pitz, G. F., Reinhold, H., & Geller, E. S. Strategies of information seeking in deferred decision making, Organizational Behavior and Human Performance, 1969.

Platt, W. Strategic intelligence production. New York: Praeger, 1957.

Polya, G. How to solve it. Princeton, New Jersey: Princeton University Press, (2nd edition, Anchor Press), 1957.

Polyani, M. Experience and the perception pattern. In K. M. Sayre & F. J. Crossman (Eds.), The modeling of mind: Computers and intelligence. Notre Dame: University of Notre Dame Press, 1963.

Pruitt, P. G. Informational requirements in making decisions. American Journal of Psychology, 1961, 74, 433-439.

Raiffa, H. Decision analysis. Massachusetts: Addison-Wesley, 1968.

Raiffa, H. Preferences for multi-attributed alternatives. Rand Corporation Memo RM-5868-DOT-RC, April, 1969.

Raiffa, H. & Schlaifer, R. Applied statistical decision theory. Boston: Harvard University, 1961.

Rapoport, A. Flights, games and debates. Ann Arbor: University of Michigan Press, 1960.

Rapoport, A. Sequential decision-making in a computer-controlled task. Mathematical Psychology., 1964, 1, 351-374.

Rapoport, A., & Wallsten, T. S. Individual decision behavior. Annual Review of Psychology, 1972, 23, 131-176.

Reingen, P. H. Risk-taking by individuals and informal groups with the use of industrial product purchasing situations as stimuli. Journal of Psychology, 1973, 85, 339-345.

Rescher, N. Delphi and values. Rand Corp. Memo #P-4182, Sept., 1969.

Rice, D. C., & Meckley, R. F. Supervision and decision-making skills in vocational education: A training program utilizing simulation techniques. Office of Education (DHEW), Bureau of Research, Washington, D. C., Final Report No. RS-51, Marcy, 1970, ED 038501.

Rigney, J. W., & DeBow, C. H. Decision strategies in AAW: I. Analysis of air threat judgments and weapons assignments. USN ONR Technical Report No. 47.

Rigby, L. V., & Swain, A. D. In-Flight target reporting--How much is "a bunch"? Human Factors, 1971, 13, 177-182.

Roberts, F. S. What if utility functions do not exist? Rand Corp. memo no. R-528-ARPA, 1970.

Roby, T. B. Belief states, evidence and action. In D. P. Hunt & D. L. Zink (Eds.), Predecisional processes in decision making: Proceedings of a Symposium, USAF AMRL-TDR-64-77, 1964, 27-46.

Roby, T. B. Belief states and the uses of evidence. Behavioral Science, 1965, 10, 255-270.

Rochester, S. R. & Gill, J. Production of complex sentences in monologues and dialogues. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 203,210.

Rousseau, W. F., & Zamora, R. M. A user's manual for the generation and analysis of probability and decision trees. SRI Final Report, Project 464531-307, 1972.

Russell, B. The axiom of infinity. Hibbert J., 1904, 809-812.

Samet, M. G. Effects of the logarithmic, quadratic and spherical payoff functions on response behavior. Doctoral Dissertation, Tufts University, 1971.

Samet, M. Quantitative interpretation of two qualitive scales used to rate military intelligence. Human Factors, 1975a, 17, 76-86.

Samet, M. Subjective interpretation of reliability and accuracy scales for evaluating military intelligence. U.S. Army Research Institute for the Behavioral and Social Sciences Technical Paper No. 260, 1975b.

Samuelson, P. A. Foundations of economic analysis. Massachusetts: Harvard University Press, 1947.

Savage, L. J. Foundation of statistics. New York: Wiley, 1954.

Savage, L. J. Elicitation of personal probabilities and expectations. Journal of the American Statistical Association, 1971, 66, 783-801.

Scalzi, J. B. Can decision making be taught? Civil Engineering - ASCE, 1970, 40, 37-38.

Schiller, F. C. S. Logic for use. Harcourt, Brace, 1930.

Schlaifer, R. Probability and statistics for business decisions. New York: McGraw-Hill, 1959.

Schlaifer, R. Analysis of decisions under uncertainity. New York: McGraw-Hill, 1969.

Schopenhauer, A. The art of controversy. In J. B. Saunders (Ed.), The Essays of Arthur Schopenhauer, New York: Wiley Book Co., no date.

Schrenk, L. P. Objective difficulty and input history in sequential decision making. Human Factors, 1964, 9, 49-55.

Schrenk, L. P. Aiding the decision maker--A decision process model. Ergonomics, 1969, 12, 543-557.

Schrenk, L. P., & Kanarick, A. F. Diagnostic decision making in a Bayesian framework. Proceedings of the 5th Annual Symposium on Human Factors in Electronics, 1967.

Schum, D. A., & DuCharme, W. M. Comments on the relationship between the impact and the reliability of evidence. Organizational Behavior and Human Performance, 1971, 6, 111-131.

Schum, D. A., DuCharme, W. M. & DePitts, K. E. Research on human multi-stage probabilistic inference processes. Organizational Behavior and Human Performance, 1973, 10, 318-348.

Schum, D. A., Goldstein, I. L., Howell, W. C., & Southard, J. F. Subjective probability revisions under several cost payoff arrangements. Organizational Behavior & Human Performance, 1967, 2, 84-104.

Schum, D. A., Goldstein, I. L., & Southard, J. F. Research on a simulated Bayesian information processing system. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 37-48.

Schum, D. A., Southard, J. F., & Wombolt, L. F. Aided human processing of inconclusive evidence in diagnostic systems: A summary of experimental evaluations. AMRL-TR-69-11, Aeromedical Research Laboratories, Wright Patterson, AFB, 1969.

Sebestyen, G. S. Decision making processes in pattern recognition. New York: Macmillan, 1962.

Seigel, J. H., & Farrell, E. J. A computer simulation model to study the clinical observability of ventilation and perfusion abnormalities in human shock states. Surgery, 1973.

Shackle, G. L. S. Time in economics. Amsterdam: North Holland Publishing Co., 1958; second printing 1967.

Sheridan, A. J., & Carlson, R. E. Decision-making in a performance appraisal situation. Personnel Psychology, 1972, 25, 339-359.

Shuford, E.H. A computer-based system for aiding decision making. In J. Spiegel & D. Walker (Eds.), Information system sciences, Proceedings of the Second Congress. Washington, D.C.: Spartan Books, Inc., 1965, 157-168.

Shuford, E. H. Cybernetic testing. Electronic Systems Division, U.S.A.F. Report No. ESD-TR-67-229, 1967.

Shuford, E.H., Albert, A., & Massengill, H.E. Admissible probability measurement procedures. Psychometrika, 1966, 31, 125-147.

Sidorsky, R.C. Some determinants of computer-aided decision effectiveness. Proceedings of 80th Annual Convention, American Psychological Association, 1972, 697-698.

Sidorsky, R. C., & Houseman, J. F. Research on generalized skills related to tactical decision making. U.S. Navy Training Device Center Report NAVTRADEVCEN TR 1329-2, 1966.

Sidorsky, R. C., Houseman, J. F., & Ferguson, D. E. Behavioral and operational aspects of tactical decision making in AAW and ASW. NAVTRADEVCEN 1329-1, 1964.

Sidorsky, R. C., & Mara, T. D. Training aspects of computer-aided decision making: I. Comparison of two approaches to man-computer interaction. NAVTRADEVCEN 1329-3, July 1968.

Siegel, J. H. & Strom, B. L. The Computer as a 'living textbook' applied to the care of the critically injured patient. Journal of Trauma, 1972, 12, 739-755.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. Psychometrika, 1959, 24, 303-316.

Simon, H. A behavioral model of rational choice. Quarterly Journal of Economics, 1955, 69, 99-118.

Slovic, P. Value as a determiner of subjective probability. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7-22-28.

Slovic, P. Influence of response mode upon the relative importance of probabilities and payoffs in risk taking. Proceedings of the 75th Annual Convention, American Psychological Association, 1967, 33-34.

Slovic, P. Manipulating the attractiveness of a gamble without changing its expected value. Journal of Experimental Psychology, 1969, 79, 139-145.

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Snapper, K. J., & Fryback, D. G. Inference based on unreliable reports. Journal of Experimental Psychology, 1971, 87, 401-404.

Snapper, K. J., & Peterson, C. R. Information seeking and data diagnosticity. Journal of Experimental Psychology, 1971, 87, 429-433.

Soelberg, P. First generalizable decision process model. MIT: Sloan School, 1966.

Soelberg, P. O. Unprogrammed decision making. Industrial Management Review, 1967, 8, 19-29.

Southard, J. F., Schum, D. A., & Briggs, G. E. Application of Bayes theorem as a hypothesis--Selection aid in a complex information processing system. Ohio State University, AMRL-TDR-64-51, 1964a.

Southard, J. F., Schum, D. A., & Briggs, G. E. Subject control over a Bayesian hypothesis-selection aid in a complex information processing system. Ohio University, AMRL-TR-64-95, 1964b.

Spruance, R. A. Foreword in Fuchida, M., & Okumiya, M. MIDWAY: The Japanese Navy's story of the battle that doomed Japan. U.S. Naval Institute: Ballantine Books, 1955.

Steiger, J. H.,. & Gettys, C. F. Best-guess errors in multi-stage inference. Journal of Experimental Psychology, 1972, 92, 1-7.

Stewart, N. R., & Winborn, B. B. A model of decision making in systematic counseling. Educational Technology, 1973, 13, 13-15.

Summers, J. K., & Bennett, E. AESOP--A final report: A prototype on-line interactive information control system. In D. E. Walker (Ed.), Information system science and technology, Washington, D.C.: Thompson Book Co., 1967, 349-357.

Swets, J. A., & Birdsall, T. G. Deferred decision in human signal detection: A preliminary experiment. Perception & Psychophysics, 1967, 2, 15-18.

Szafran, J. The effects of ageing on professional pilots. In Price, J. H. (Ed.), Modern trends in psychological medicine II, New York: Appleton-Century-Crofts, 1970.

Tanner, W. P., Jr. A theory of recognition. Journal of the Acoustical Society of America, 1956, 28, 882-888.

Thrall, R. M., Coombs, C. H., & Davis, R. L. Decision processes. John Wiley & Sons, 1954.

Toda, M. Measurement of subjective probability distributions. Decision Sciences Laboratory Technical Documentary Report No. ESD-TDR-63-407, July 1963.

Triggs, T. J. Decision flow diagrams in information management systems. BBN Report 2617. 1973, Submitted to U.S. Army Research Institute for the Behavioral and Social Sciences under contract DAHC-19-72-C-0019.

Tuddenham, W. J. The use of logical flow charts as an aid in teaching roentgen diagnosis. American Journal of Roentgenology, 1968, 102, 797, 803.

Tune, G. S. Response preferences: A review of some relevant literature. Psychological Bulletin, 1964, 61, 286-302.

Tversky, A., & Kahneman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.

Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973, 5, 207-232.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Van Naerssen, R. F. A scale for the measurement of subjective probability. Acta Psychologica, 1962, 20, 159-166.

Vlek, C. A. Multiple probability learning: Associating events with their probabilities of occurrence. Acta Psychologica, 1970, 33, 207-232.

Von Neumann, J. & Morgenstern, O. Theory of games and economic behavior, 2nd ed., New Jersey: Princeton University Press, 1947.

Wagenaar, W. A. Subjective randomness and the capacity to generate information. In A. F. Sanders (Ed.), Attention and performance III, Acta Psychologica, 1970, 33, 233-242.

Wald, A. Sequential analysis. New York: Wiley, 1947.

Wald, A. Statistical decision functions. New York: Wiley, 1950.

Wallas, G. The art of thought. London: Jonathan Cape, 1925.

Wason, P. The drafting of rules. New Law Journal, 1968, 118, 548-549.

Wason, P. The psychology of deceptive problems. *New Scientist*, 1974, *63*, 382-385.

Whitehead, S. F., & Castleman, P. A. Evaluation of an automated medical history in medical practice. *Proceedings of the 1974 Conference of the Society for Computer Medicine* (to be published).

Williams, A. C., Jr. & Hopkins, C. O. Aspects of pilot decision-making. WADC TEchnical Report 58-522, 1958.

Winkler, R.L., & Murphy, A.H. Evaluation of subjective precipitation probability forecasts. In *Proceedings of the First National Conference on Statistical Meteorology*, American Meteorology Society, Boston, 1968, 148-157.

Wright, P. Writing to be understood: Why use sentences? *Applied Ergonomics*, 1971, *2*, 207-209..

Yeh, S.Y., Betyar, L., & Hon, E.H. Computer diagnosis of fetal heart rate patterns. *American Journal of Obstetrics and Gynecology*, 1972, *114*, 890-897.

Yntema, D. B., & Klem, L. Telling a computer how to evaluate multidimensional situations. *IEEE Transactions on Human Factors Engineering*, 1965, 3-13.

Yntema, D. B. & Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. *IRE Transactions on Human Factors in Electronics*, 1961, 20-26.

Zadeh, L. A. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics*, 1973, *SMC-3*, 28-44.

HUMAN FACTORS DISTRIBUTION LIST

NOTE

Mailing labels are prepared, when needed, as a computer listing, the source of which is updated on a weekly basis. It is not practical to prepare distribution lists each time labels are prepared. Therefore such lists are prepared semiannually, and a slight discrepancy may exist between the addressees on this list and those appearing on the labels used to distribute this publication.

Chief of Naval Training
Attn: Captain A. E. McMichael, N-3
Pensacola, FL 32508

Chief of Naval Training
Attn: Captain B. C. Stone
Pensacola, FL 32508

Chief of Naval Training
Attn: Dr. W. Maloy, Code 01A
Pensacola, FL 32508

Chief of Naval Material
Attn: Mr. A. L. Rubinstein, MAT-03424
Navy Department
Washington DC 20360

Commanding Officer
Naval Submarine Base, New London
Attn: Psychology Section
Box 00
Groton, CT 06340

Chief of Naval Air Training
Attn: Joseph L. Ulatoski
Naval Air Station
Corpus Christi, TX 78419

Commander Training Command
Attn: Educational Advisor
U.S. Pacific Fleet
San Diego, CA 92147

Chief Naval Research
Attn: Code 458 Navy Department
Arlington, VA 22217

COMDT of Marine Corps
Code A03C
Washington, DC 20380

Air Force Human Resources Lab/DOJZ
Brooks Air Force Base
Texas 78235

CDR NAVAIR DEVEL CTR
Attn: Human Engrg Br
Warminister, PA 18974

Commander
Naval Electronics Systems Command
Code 03
Washington, DC 20360

NAVAERSPMEDINST
NAVAEROSPREGMEDCEN
ATTN: Ch Avia Psych Div
Pensacola, FL 32512

Chief Naval Operations
Attn: CDR H. J. Connery
OP-987 M42 Navy Dept
Washington, DC 20350

HUMAN FACTORS DISTRIBUTION LIST

Commander Training Command
Attn: Educational Advisor
U. S. Atlantic Fleet
Norfolk, VA 23511

Assistant Secretary of the Navy (R&D)
Attn: Dr. Samuel Koslov, 4E741
Navy Department
Washington, DC 20350

Director of Defense Research & Engineering
Attn: Lt Col Henry Taylor, OAD (R&D)
Washington, DC 20301

Dir Human Res Rsch Org
300 N Washington St
Alexandria, VA 22314

Commander
NAVAIRSYSCOM
Code 03
Washington, DC 20360

CDR NAVSHIPS SYS CMD
NAVSHIPS SYS CMD HQS
SHIPS 03H Geo. N. Graine
Washington, DC 20360

Chief Naval Operations
Attn: Dr. J. J. Collins
OP-987F Navy Dept
Washington, DC 20350

ERIC Clearinghouse UN
EDUCAT MEDIA--TECH
Stanford Univ
Stanford, CA 94305

Bureau Naval Personnel
Attn: PERS A3
Arlington Annex
Washington, DC 20370

Chief Naval Research
Psychological Sciences
Code 450 Navy Dept
Arlington, VA 22217

Chief Naval Material
MAT 031M
Washington, DC 20360

USAF Human Res Lab
Personnel Rsch Div
Lackland AFB, TX 78236

Human Resources Resch
Orgztn Div 6 Aviation
P. O. Box 428
Fort Rucker, AL 36360

National Science Foundation
Attn: Dr. Henry S. Odbert
1800 G St NW
Washington, DC 20550

Chief Naval Operations
Attn: M. K. Malehorn
OP-14C Navy Dept
Washington, DC 20350

USAF Human Res Lab
AFHRL-FT
Flying Trng Div
William AFB, AZ 85224

USAF Human Res Lab
AFHRL-TT
Tech Tng Div
Lowry AFB, CO 80230

Commander
NAVORDSYSCMD
Code 03
Washington, DC 20360

HUMAN FACTORS DISTRIBUTION LIST

USAF Human Res Lab
AFHRL-AS
Advance Systems Div
Wright-Patterson AFB, OH 45433

U.S. Army Research Institute
Research Lab
Commonwealth Bldg (Rm 239)
1320 Wilson Blvd
Arlington, VA 22209

Commanding Officer
PERS-TRNG RESH DEV LAB
San Diego, CA 92152

CNETS
Code N-241
NAS Pensacola, FL 32508

Naval Education & Training Support
Center (Pacific)
Code N1
Fleet Station Post Office Bldg
San Diego, Ca 92132

Mr. Sam Campbell
Grumman Aerospace Corporation
Plant 47
Bethpage, L. I., New York 11714

Mr. Robert E. Coward
Chief, Instructional Technology Div
ADC/DOTI
Ent AFB CO 80912

Commanding Officer 88
Naval Training Equipment Center
Orlando, FL 32813

Defense Documentation Center 12
Cameron Station
Alexandria, VA 22314

NAV PERSONNEL RESCH AND
DEVELOPMENT LABORATORY
Attn: Library
Bldg 200 Rm 3307 WNY
Washington, DC 20390

CNETS
Code N-2, Bldg 45
(Dr. Charles Havens)
NAS Pensacola, FL 32508

Dr. Mark A. Hofmann
U.S. Army Aeromedical Research Lab
P.O. Box 577
Fort Rucker, Alabama 36360

Mr. Harold Kottmann
ASD/SMSE
Wright-Patterson Air Force Base
Ohio 45433

Mr. Arthur Doty
Wright Patterson AFB (ENCT)
Dayton, Ohio

COMNAVAIRPAC
Box 1210 USNAS
ATTN: Code 316
North Island
San Diego, CA 92135

The Field Artillery School
Target Acquisition Department
(Mr. Eugene C. Rogers)
Ft. Sill, Oklahoma 73503

USAF Human Res Lab
AFHRL/OR
Occupational & Manpower Research Div
Lackland AFB TX 78236

HUMAN FACTORS DISTRIBUTION LIST

HQS AF SYS CMD DLSL
Ofc Scientific Rsch
Andrews AFB
Washington, DC 20331

Commander
NAVSUPSYSCMD
Code 03
Washington, DC 20360

USAF Human Res Lab
AFHRL/SM
Computational Sciences Div
Lackland AFB TX 78235

Human Res Rsch Org
Div No. 1 Sys Oper
200 N. Washington ST
Alexandria, VA 22314

CO NAV MISSILE CTR
Attn: Hd Human Factors
Engrg Br
Point Mugu, CA 93042

Commanding Officer
Navy Medical
Neuropsychiatric Resch Unit
San Diego, CA 92152

CO NAVAIR TECH TRNG
NAS Memphis
Attn: Dr. G. D. Mayo
Hd Research Br
Millington, TN 38054

Dir DEF RSCH-ENGRG
ARPA
Behavioral Science Div
Attn: LCOL A. W. Kibler
Washington, DC 20301

Scientific Technical
Information Office
NASA
Washington, DC 20546

CH RSCH OFC
Ofc Dep Ch of Staff for Pers
Dept of Army
Washington, DC 20310

Chief of Naval Technical Training
NAS Memphis 75
Attn: Code 34 Dr. Harding
Millington, TN 38054

Dr. John Meyer
HDQTRS Air Training Command
XPT
Randolph AFB, TX 78148

Joseph J. Cowan
Ch PERS RSCH BRANCH
USCG HQ PO-1 STA 3-12
400 Seventh St. S. W.
Washington, DC 20590

Executive Editor
Psycho Abstracts
American Psych Assoc
1200 17th St NW
Washington, DC 20036

Dr. Ralph R. Canter
Dir MIL MANPWR RSCH
OSAD M-RA MR-U
Pentagon RM 3D960
Washington, DC 20301

Dr. John W. Weisz
Dir Human Engrg Lab
USA Aberdeen Rsch Devel Ctr
Aberdeen Proving Grounds
MD 21005